National Library
of Canada

Bibliothèque nationale
du Canada

Canadian Theses Service

Service des thèses canadiennes

Ottawa, Canada
K1A 0N4

NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Reproduction in full or in part of this microform is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30, and subsequent amendments.

AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30, et ses amendements subséquents.

NL-339 (r. 88/04) c

Canada

UNIVERSITY OF ALBERTA

THEORY OF MIND AND AUTISTIC CHILDREN: TEACHING THE APPEARANCE-REALITY DISTINCTION

BY

ELIZABETH MARGARET STARR

A THESIS SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND
RESEARCH IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE
DEGREE OF DOCTOR OF PHILOSOPHY

IN

SPECIAL EDUCATION

DEPARTMENT OF EDUCATIONAL PSYCHOLOGY

EDMONTON, ALBERTA
SPRING, 1992

National Library
of Canada

Bibliothèque nationale
du Canada

Canadian Theses Service

Service des thèses canadiennes

Ottawa, Canada
K1A 0N4

The author has granted an irrevocable non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of his/her thesis by any means and in any form or format, making this thesis available to interested persons.

The author retains ownership of the copyright in his/her thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without his/her permission.

L'auteur a accordé une licence irrévocable et non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de sa thèse de quelque manière et sous quelque forme que ce soit pour mettre des exemplaires de cette thèse à la disposition des personnes intéressées.

L'auteur conserve la propriété du droit d'auteur qui protège sa thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

ISBN 0-315-73037-4

Canada

UNIVERSITY OF ALBERTA
RELEASE FORM

NAME OF AUTHOR: ELIZABETH MARGARET STARR

TITLE OF THESIS: THEORY OF MIND AND AUTISTIC CHILDREN: TEACHING THE APPEARANCE-REALITY DISTINCTION

DEGREE: DOCTOR OF PHILOSOPHY

YEAR THIS DEGREE GRANTED: SPRING 1992

Permission is hereby granted to the University of Alberta Library to reproduce single copies of this dissertation and to lend or sell such copies for private, scholarly or scientific research purposes only.

The author reserves other publication rights, and neither the dissertation nor extensive extracts from it may be printed or otherwise reproduced without the author's written permission.

Elizabeth Starr
12 Foxbar Road
Toronto, Ontario
M4V 2G6

Date: March 27/92

UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

The undersigned certify they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled **THEORY OF MIND AND AUTISTIC CHILDREN: TEACHING THE APPEARANCE-REALITY DISTINCTION** submitted by Elizabeth Margaret Starr in partial fulfillment of the requirements for the degree of **DOCTOR OF PHILOSOPHY** in **SPECIAL EDUCATION**.

Dr. D. Baine (Supervisor)

Dr. R. Mulcahy (Supervisor)

Dr. G. Bisanz

Fern Snart

Dr. F. Snart

L. Wilgosh

Dr. L. Wilgosh

Dr. L. McDonald

Dr. C. Crealock

Date: March 12/92

ABSTRACT

Theory of mind refers to an individual's ability to recognize and understand the existence of mental states in oneself and others, and to realize that others can hold conceptual perspectives that differ from one's own. The precursors of this ability begin to emerge in the second year of life in non-disabled children, in the form of symbolic play, perceptual perspective taking, and joint-attentional gestures, and appear to be fully developed by about seven years of age. Research has shown, however, that children with autism fail to develop a fully functional theory of mind, a deficit that is considered to be autism-specific, and considered by some researchers to be the origin of autistic children's well documented cognitive and social deficits.

The purpose of this study was to investigate the possibility of using a behavioral approach to teach one aspect of theory of mind, the appearance-reality (A-R) distinction, to children diagnosed as autistic. Using a pre-experimental case study approach, Direct Instruction (DI) was used to teach Color and Size appearance-reality distinction tasks.

A total of five subjects ranging in age from 9.4 to 12.1 were included in the study. Pre- and post-testing measures included a Color task, a Size task, and an Object Identity task (for generalization purposes). Instruction was provided on three Color tasks and three Size tasks until criterion had been reached (three out of three tasks correct in each domain), or until 10 instructional sessions had been completed in each domain. Probes, in which performance on these tasks was tested, were conducted approximately twice per week. Two additional Color and Size tasks were used as well, as measures of generalization.

Of the five subjects included in the study, three demonstrated mastery of the Color appearance-reality distinction tasks after intervention. One subject appeared to gain mastery of the tasks through repetition of the baseline tasks, and one subject failed to achieve mastery on the Color A-R tasks. Three subjects attained criterion on the Size A-R tasks as well. None of the subjects seemed to understand the requirements of the Object Identity tasks, and were unable to answer the control questions appropriately on these tasks.

The investigator concluded that the Direct Instruction procedure, although allowing some subjects to reach criterion, encouraged other subjects to demonstrate their mastery of the task instructions without understanding the underlying concept. It appeared that for some subjects, the linguistic complexity of the task presentation illuminated their semantic and pragmatic language difficulties rather than their implicit understanding of the appearance-reality distinction concept. Suggestions for future research included a call for more comprehensive language assessment than has been conducted to date if children with autism are to be included as research subjects.

ACKNOWLEDGEMENT

The acknowledgement page, while marking the end of a odyssey and a joy to write from that perspective, is also perhaps one of the most difficult pages to write, for fear of omitting the contribution of an especially important person, or of becoming maudlin. This page is written with the hope that both *faux pas* are avoided.

As in all dissertations, many individuals and organizations deserve special recognition for their support of my effort in this endeavour, or for their significant contribution that enabled its completion. First of all, my deepest gratitude is extended to my co-supervisors, Dr. David Baine and Dr. Bob Mulcahy for their endless hours of consultation, enduring patience, and well considered suggestions. Thank you for at least *appearing* interested in talking with me when you saw me coming yet again! My appreciation is also extended to my committee, Dr. Lorraine Wilgosh, Dr. Fern Snart, Dr. Gay Bisanz, and Dr. Linda McDonald, who didn't seem to balk too much at the bulk with which they were presented, and who provided four additional perspectives. Thanks also to Dr. Carol Crealock of the University of Western Ontario who served as external reader.

My heartiest thanks is extended to the following individuals and organizations who contributed to the study by providing either procedural, administrative, or financial assistance: the Edmonton Autism Society, for their help in recruiting subjects; the Edmonton Public School Board, for giving me access to their system; principals Mr. Keith Parker and Mr. Larry Taylor, for allowing me to conduct my research in their schools; the various teachers, students, and parents concerned, for their cooperation; Dr. Gary Holdgrafer, for his advice in test selection; the Glenrose Rehabilitation Hospital, Tony McClellan, and Edmonton Public, for their loan of the Leiter International Performance Scale; Dr. Joanne Volden, for her assistance in obtaining language samples; Frank Symons, for inter-rater reliability services; Bev Baker-Hofmann, for her help in devising the instructional method; Shawn Crawford, for explaining Cricketgraph intricacies; Dr. Simon Baron-Cohen of the Institute of Psychiatry, London, England, who took an interest in my project and provided valuable articles; and finally, the Social Sciences and Humanities Research Council, the University of Alberta, the G. Allan Roeher Institute, and the Scottish Rite Foundation for their support of my research, and for rescuing me from imminent penury.

I must also thank a few special friends whose belief in my ability to complete this opus didn't waver even when mine did, and who gave me much needed distractions: Merna Schmidt; Rosalind Drodge; David and Shelley Veats; Deb and Bruce Brandell; Joyce Maycher; Bev Baker-Hofmann and Greg Hofmann; Sara Yarwood; Dr. Rod Beattie, my

office-mate who charged cheap rent; Dr. Jerry Rose, my former office-mate who charged no rent; the Ed Psycho Drama Players, who provided just enough lunacy for me to keep my sanity; and the Richard Eaton Singers and Da Camera Singers, who took care of the musical side of my life.

Finally, my deepest thanks and gratitude is reserved for my family who has encouraged and supported me in so many ways over the years. My mother's attentiveness in listening to my thesis ravings even when she did not have the faintest idea what I was talking about was much appreciated!

TABLE OF CONTENTS

CHAPTER PAGE

1. INTRODUCTION AND RATIONALE 1
2. REVIEW OF THE LITERATURE 5
    - The Autistic Syndrome 5
    - Theory of Mind, its Precursers and Development 7
        - Symbolic Play 9
    - Theory of Mind Research with Autistic Children 13
        - False-Belief Studies 13
        - Causes of Emotion 22
        - Appearance–Reality Distinction Studies 24
    - Appearance-Reality Research with Non-Autistic Children 26
    - Instruction of Autistic Children 30
        - Space 31
        - Time 31
        - Duration 31
    - Direct Instruction 32
    - Concept Generalization 33
    - Direct Instruction and Autistic Children 34
    - Summary and Limitations of Current Research 36
3. METHODOLOGY 41
    - Inclusion Criteria 41
    - Subjects 41
    - Setting 43
    - Inclusion Materials and Procedure 43
        - Memory Test 43
        - Appearance-Reality Inclusion Tasks 44
            - Color 44
            - Size 45
            - Object Identity 46
        - Peabody Picture Vocabulary Test–Revised 47
        - Leiter International Performance Scale 48
        - Language Sample 49
        - Scoring 50

Instructional Materials and Procedure 50
  Materials 50
  Procedure 51
  "Bigger" Tasks 56
  Scoring 56
Experimental Design 57
  Data Collection 58
  Alternate Question Format 59
  Videotaping 59
Inter-rater Reliability 59
Ethical Considerations 60
4. RESULTS 62
Individual Subject Data 62
  Nicky 64
    Description 64
    Inter-Rater Reliability 66
      Color tasks 66
      Size tasks 67
    Results 68
      Color tasks 68
      Size tasks 73
      Intrasubject comparison on the Color and Size tasks 77
      Object Identity tasks 80
    Summary 81
  David 82
    Description 82
    Inter-Rater Reliability 83
      Color tasks 83
      Size tasks 83
    Results 84
      Color tasks 84
      Size tasks 88
      Intrasubject comparison on the Color and Size tasks 92
      Object Identity tasks 95
    Summary 97

Brian 98
  Description 98
  Inter-Rater Reliability 100
    Color tasks 100
    Size tasks 100
  Results 101
    Color tasks 101
    Size tasks 104
    Intrasubject comparison of Color and Size tasks 107
    Object Identity tasks 110
  Summary 111
Bruce 112
  Description 112
  Inter-Rater Reliability 113
    Color tasks 113
    Size tasks 113
  Results 114
    Color tasks 114
    Size tasks 116
    Intrasubject comparison on the Color and Size tasks 122
  Summary 126
Steve 127
  Description 127
  Inter-Rater Reliability 128
    Color tasks 128
    Size tasks 128
  Results 129
    Color tasks 129
    Size tasks 133
    Intrasubject comparison on the Color and Size tasks 135
    Object Identity tasks 138
  Summary 139
Summary of Results 140

5. DISCUSSION 142
  Post-test Data 142
  Color and Size Appearance-Reality Task Performance 143
    Examination of Behavior and Error Patterns 147
  Object Identity Task Performance 150
  Evaluation of Instructional Methodology 151
    Effectiveness of Direct Instruction Procedure 151
    Effectiveness of Task Presentation Method 152
    Effectiveness of Reinforcement Procedure 153
    Effectiveness of Question Format 153
  Limitations of the Current Study and 155
    The Role of Language 155
    Relationship of Success on Appearance-Reality Tasks to 156
    Symbolic Play and Theory of Mind 157
  Summary and Conclusions 158
REFERENCES 161
APPENDIX A. THEORY OF MIND STUDY SUMMARIES 170
APPENDIX B. SCORING FORMS 177
APPENDIX C. INSTRUCTION SCRIPTS FOR COLOR AND SIZE TASK 182
APPENDIX D. INSTRUCTIONAL FORMAT ANALYSIS 189
APPENDIX E. INFORMATION AND CONSENT FORMS 191
APPENDIX F. SESSION NUMBERS AND DATES 196

LIST OF TABLES

Table 3.1. Subject Background Data 42
Table 4.1. Nicky—Inter-Rater Reliability on Color Tasks 67
Table 4.2. Nicky—Inter-Rater Reliability on Size Tasks 68
Table 4.3. Nicky—Data Breakdown of Color Sessions When Alternate Question Format Used 70
Table 4.4. Nicky—Data Breakdown of Size Sessions When Alternate Question Format Used 75
Table 4.5. David—Inter-Rater Reliability on Color Tasks 83
Table 4.6. David—Inter-Rater Reliability on Size Tasks 84
Table 4.7. David—Data Breakdown of Color Sessions When Alternate Question Format Used 86
Table 4.8. David—Data Breakdown of Size Sessions When Alternate Question Format Used 91
Table 4.9. Brian—Inter-Rater Reliability on Color Tasks 100
Table 4.10. Brian—Inter-Rater Reliability on Size Tasks 101
Table 4.11. Brian—Data Breakdown of Color Sessions When Alternate Question Format Used 103
Table 4.12. Brian—Data Breakdown of Size Sessions When Alternate Question Format Used 106
Table 4.13. Bruce—Inter-Rater Reliability on Color Tasks 114
Table 4.14. Bruce—Inter-Rater Reliability on Size Tasks 114
Table 4.15. Bruce—Data Breakdown of Color Sessions When Alternate Question Format Used 117
Table 4.16. Bruce—Data Breakdown of Size Sessions When Alternate Question Format Used 120
Table 4.17. Steve—Inter-Rater Reliability on Color Tasks 128
Table 4.18. Steve—Inter-Rater Reliability on Size Tasks 129
Table 4.19. Steve—Data Breakdown of Color Sessions When Alternate Question Format Used 131
Table 4.20. Steve—Data Breakdown of Size Sessions When Alternate Question Format Used 135
Table 5.1. Subject Post-Test Performance 143

LIST OF FIGURES

Figure 3.1. Size task apparatus. 45

Figure 3.2. Color instructional tasks. 52

Figure 3.3. Size instructional tasks 53

Figure 4.1. Nicky—Number of correct responses on tests of Color tasks across conditions. 69

Figure 4.2. Nicky—Phase 4 intervention on Color tasks. 69

Figure 4.3. Nicky—Number of correct responses on tests of Size tasks across conditions. 74

Figure 4.4. Nicky—Phase 4 intervention on Size tasks. 74

Figure 4.5. Nicky—Comparison of Phase 4 intervention on Color and Size tasks 78

Figure 4.6. Nicky—Comparison of performance on tests of Color and Size tasks 79

Figure 4.7. David—Number of correct responses on tests of Color tasks across conditions. 85

Figure 4.8. David—Phase 4 intervention on Color tasks. 85

Figure 4.9. David—Number of correct responses on tests of Size tasks across conditions. 90

Figure 4.10. David—Phase 4 intervention on Size tasks. 90

Figure 4.11. David—Comparison of Phase 4 intervention on Color and Size tasks 93

Figure 4.12. David—Comparison of performance on tests of Color and Size tasks 94

Figure 4.13. Brian—Number of correct responses on tests of Color tasks across conditions. 102

Figure 4.14. Brian—Phase 4 intervention on Color tasks. 102

Figure 4.15. Brian—Number of correct responses on tests of Size tasks across conditions. 105

Figure 4.16. Brian—Phase 4 intervention on Size tasks. 105

Figure 4.17. Brian—Comparison of Phase 4 intervention on Color and Size tasks 108

Figure 4.18. Brian—Comparison of performance on tests of Color and Size tasks 109

Figure 4.19. Bruce—Number of correct responses on tests of Color tasks across conditions. 115

Figure 4.20. Bruce—Number of correct responses on tests of Size tasks across conditions. 118
Figure 4.21. Bruce—Phase 4 intervention on Size tasks. 118
Figure 4.22. Bruce—Comparison of performance on tests of Color and Size tasks 123
Figure 4.23. Steve—Number of correct responses on tests of Color tasks across conditions. 130
Figure 4.24. Steve—Phase 4 intervention on Color tasks. 130
Figure 4.25. Steve—Number of correct responses on tests of Size tasks across conditions. 134
Figure 4.26. Steve—Phase 4 intervention on Size tasks. 134
Figure 4.27. Steve—Comparison of Phase 4 intervention on Color and Size tasks 136
Figure 4.28. Steve—Comparison of performance on tests of Color and Size tasks 137
Figure 5.1. Intersubject comparison on Color tasks. 145
Figure 5.2. Intersubject comparison on Size tasks. 146

CHAPTER ONE

INTRODUCTION AND RATIONALE

As children develop, they begin to acquire an understanding not only of the physical world, but also of the mental world. They begin to recognize the existence of mental states and to understand beliefs, intentions and desires in both themselves and others (Olson, 1988; Olson, Astington, & Harris, 1988; Perner, Frith, Leslie, & Leekam, 1989; Wellman, 1985). This ability, in turn, enables children to understand and predict the behavior of others (Leslie & Frith, 1988). This understanding has been termed "theory of mind" (Premack & Woodruff, 1978), and it involves the ability to form meta-representations, or representations of representational states (Leslie, 1987; Olson et al., 1988).

The development of a theory of mind is crucial for the development of appropriate and meaningful social interaction (Baron-Cohen, Leslie, & Frith, 1985; Perner, 1988) and intentional communication (Bretherton & Beeghly, 1982). As Wellman (1985) observed, for a being who possessed no ability to understand the mental activities of others:

> The social world, the world of self and others would be an impoverished place...All those aspects of behavior that we attribute to the mind would be missing from its perspective. Persons would be seen and heard but there would be no notion of a backlog of ideas or beliefs organizing their actions and personalities. Indeed, for this hypothetical being, no one could be construed as possessing private persona; public present behavior would have no deeper meaning. The concept of a lie would be inconceivable, as would any concept that depends on an awareness of the independent or quasi-determinate relationship between internal states and external events: notions such as illusions, beliefs, hunches, mistakes, guesses, or deceptions. (p. 170)

Without this skill, the ability to explain and predict the behavior of others, or understand why people do certain things would be absent (Baron-Cohen et al., 1985, 1986; Leslie, 1988; Premack & Woodruff, 1978), as would the ability to influence states of mind, a common goal of both verbal and nonverbal communication (Perner et al., 1989).

Theory of mind is an encompassing term that includes a number of more specific concepts, including false belief attribution, and appearance-reality (A-R) distinction. False belief attribution refers to the ability to understand that different people can have different beliefs about the same situation (Baron-Cohen et al., 1985). For example, a child may observe a second child place a toy in a particular location and then depart. The child may then witness a third child come and put the toy in a different location. The observing child who has acquired a theory of mind, would predict that upon returning to the room, the second child will look for the toy in its original location, even though the observing child knows that the toy is now in a different location. Thus, the observing child can attribute a false belief to the first child, taking into account the first child's limited knowledge of the situation, despite the fact that the observing child's own belief regarding the situation is different.

Appearance-reality distinction refers to the ability that young children develop that enables them to represent, on the one hand, what an object appears to be, while simultaneously representing what the object actually is. For example, a child may see what looks to be a rock. However, upon manipulating the "rock", the child discovers that in reality the object is a sponge that has been painted to look like a rock. Children are said to understand the appearance-reality distinction if they are able to state that the "rock" looks like a rock but is really a sponge, thus demonstrating their ability to understand, and mentally represent these contrasting representations. In other words, appearance-reality distinction concerns the ability to attribute mental states to oneself rather than to others (Baron-Cohen, 1989a) as does false-belief. As will be seen in the following chapter, the appearance-reality distinction also manifests itself in tasks dealing with true and apparent color, size, smell, and sound. The appearance-reality distinction can thus be demonstrated through all senses.

Flavell, Green, and Flavell (1986) cited three reasons relating to the importance of the acquisition of appearance-reality distinction. First, "the relation between appearance and reality figures importantly in everyday perceptual, conceptual, emotional, and social activity—in misperceptions, misexpectations, misunderstandings, false beliefs, deception, play, fantasy and so forth" (p. 1). Second, it is believed that the development of appearance-reality distinction is a universal acquisition in normal populations, and that it may either mediate or be a prerequisite for Piagetian conservations. A study conducted in China (Flavell, Zhang, Zou, Dong, & Qi, 1983) demonstrated that Chinese preschool children's performance was very similar to the performance of American children despite cultural differences. Thirdly, the development of appearance-reality distinction is important in that it "is part of the larger development of our conscious knowledge about our own and other minds—thus of metacognition and of social cognition" (Flavell et al., 1986, p. 2).

A study conducted with young Baka children of southeast Cameroon demonstrates the universality of another aspect of theory of mind. Avis and Harris (1991) used a belief-desire reasoning task to test the children's conception of mind. The investigators found that the Baka children responded in a similar fashion as North American and European children. That is, a majority of older children (mean age of five years) was able to predict correctly that an adult, believing that a desirable food was in a container that had, in fact, been removed to a hiding place, would feel happy rather than sad before lifting the container's cover, and sad rather than happy after discovering that the container was empty. Although a minority of younger children (mean age of three-and-a-half) was also able to do the task correctly (six out of 17), the pattern of results replicated that of North American and European studies.

Most nonhandicapped children, and even mentally handicapped children seem to develop these various aspects of theory of mind, at least in part, by about age four (Astington & Gopnik, 1991; Cox, 1991; Leslie, 1987, 1988). Recent research, though, has shown that most children diagnosed as autistic apparently fail to develop, or experience severe delays in acquiring, a theory of mind. It appears to be an autism-specific deficit (Baron-Cohen et al., 1985; Leslie, 1987, 1988) that is independent of the mental retardation commonly associated with autism. Some researchers (Baron-Cohen,1989c, 1989d; Baron-Cohen et al., 1985, 1986; Leslie, 1987, 1988; Leslie & Frith, 1988) believe that this deficit constitutes the basis of the cognitive deficit underlying the autistic syndrome, and that autistic children may indeed experience the world as Wellman's (1985) hypothetical being does.

Although a number of studies investigating theory of mind in autistic children have been conducted, all of them demonstrating that autistic children have theory of mind deficits, not one of them has included an instructional procedure, although the possibility of such intervention has been suggested (Baron-Cohen, 1989d, Leslie & Frith, 1988). For example, Leslie and Frith (1988) suggested that with maturation or practice autistic children may be able to increase their facility with mental state concepts, at least to some extent. Baron-Cohen (1990a) has speculated on the potential of a social-cognitive training program to improve the autistic child's theory of mind deficits and is in the process of implementing such a program (Baron-Cohen, personal communication).

Given the significance of acquiring a theory of mind as discussed above, and the experimental results that have repeatedly demonstrated autistic children's theory of mind deficits, it seems that efforts at remediating those deficits are of importance. At this time, it is unknown whether remediation can effect a change in autistic children's metarepresentational ability. A few theory of mind experiments with normally developing three year old children have included brief teaching paradigms (Flavell, Flavell, & Green, 1983; Flavell et al., 1986), but even those using simplified rather than standard appearance-reality tasks, have failed to elicit understanding of the appearance-reality distinction. These results suggest that this skill is one that may emerge only through maturation and is not amenable to instruction. However, maturation was also thought to account for the development of conservation skills (Ginsburg & Opper, 1979), a belief that has been disproven in more recent research in which conservation has been taught to so-called preoperational children (Braine & Shanks, 1965; Case, 1985; Cohen, 1983). Given the postulated connection between the development of appearance-reality distinction and conservation skills (Flavell et al., 1986), and the demonstrated ability to teach conservation, it is possible that previous attempts to teach appearance-reality distinction concept were either too complicated from a linguistic perspective, or of too short duration for the children, rather than being impossible for them to understand.

Even if the acquisition of a theory of mind does follow a strict developmental sequence that is unaided by instruction, the case of autism might be different. Theory of mind teaching paradigms have only been attempted with children who, according to the developmental hypothesis, were too young to have acquired a theory of mind in the first place (i.e., less than four years of age). However, studies with autistic populations have shown that these children, although of measured mental ages far beyond that at which a theory of mind has been observed in nonautistic children, either demonstrate no evidence of a theory of mind, or very restricted representational abilities. Thus, autistic children are qualitatively different, and the results of training attempts with nonautistic populations may not be generalizable to autistic populations.

The purpose of the current study, then, was to develop and implement an intervention program, with the intention of determining whether or not intensive short-term instruction using a behavioral approach would be beneficial in remediating one aspect of autistic children's theory of mind deficits, specifically, the A-R distinction. Instruction consisted of variations of the appearance-reality distinction tasks used in previous studies (Baron-Cohen, 1989a; Flavell et al., 1983, Flavell et al., 1986; Gopnik & Astington, 1988).

Appearance-reality distinction was chosen for instruction primarily because little is known about how autistic children perform on A-R tasks. Only one study (Baron-Cohen, 1989a) has examined this aspect of theory of mind in autistic children, and no instructional attempts were made.

CHAPTER TWO

REVIEW OF THE LITERATURE

This chapter begins with an overview of the autistic syndrome: its definition according to DSM-IIIR (American Psychiatric Association, 1987), epidemiology, and the behavioral features of the syndrome. Next, Leslie's (1987, 1988) theory regarding the development of theory of mind is considered, including discussion on theory of mind precursers and the significant contribution of symbolic play on the development of theory of mind. The chapter continues with an examination of the research concerning theory of mind that has been conducted with autistic children. Thus, studies dealing with first- and second-order false-belief, understanding of the causes of emotion, and studies dealing with the appearance-reality distinction in both autistic and non-autistic children are considered. The chapter concludes with an explanation of Direct Instruction and its benefits for use with autistic children.

The Autistic Syndrome

Autism, first described by Kanner in 1943 as "autistic disturbances of affective contact" (Kanner, 1985), is a Pervasive Developmental Disorder characterized by qualitative impairments in reciprocal social interaction, imaginative activity, and in the development of verbal and nonverbal communication skills (American Psychiatric Association, 1987). These deficits are referred to by Wing as the "triad of impairments" (Frith, 1989b). Although theorists in the past have believed that autism is of psychogenic origin, brought on by "cold" and rejecting parents, (Bettelheim, 1959, 1967; Tinbergen & Tinbergen, 1983), there exists overwhelming evidence that autism has biological roots (Courchesne, 1989, 1991a, 1991b; Gillberg, 1990; Rimland, 1985; Rutter, 1978, 1983). Psychogenic theories have now been discredited.

Until recently, age of onset had to be before 30 to 36 months of age for a diagnosis of autism to be made (American Psychiatric Association, 1980; Rutter, 1978). However, this age limit is not included in current APA diagnostic criteria (American Psychiatric Association, 1987; Gillberg, 1990) for a number of reasons, including the fact that in many cases there is a problem in obtaining a reliable medical history when a child is first referred for diagnosis after the age of three (Gillberg, 1990).

Early studies (Lotter, 1985; Wing, 1980) estimated the autistic syndrome to occur in approximately four to five children in every 10,000, with a male to female sex ratio of 3:1 or 4:1 (American Psychiatric Association, 1987). More recent studies, however, (Gillberg, Steffenberg, & Jakobsson, in Gillberg, 1990) have estimated the prevalence of autism to range from 6.6 to 12.4 per 10,000 children. A Canadian epidemiological study of autistic syndromes (Bryson, Clark, & Smith, 1988) found a prevalence of 10 per 10,000 children, with a male to female ratio of 2.5:1. Bryson et al. (1988) attribute these differences from previous epidemiological studies to three possible reasons: (a) the exhaustive population screening of their study; (b) new diagnostic criteria; and (c) true differences in prevalence due to time and/or place of their study.

While about 20 to 25% of autistic children evidence average or above average intellectual ability (Wulff, 1985), IQs of the majority of autistic children lie within the mentally retarded range of intellectual functioning (Gillberg, 1990; Maltz, 1981; Rutter, 1985, Sigman, Ungerer, Mundy, & Sherman, 1987). However, unlike mentally retarded children, autistic children display a wide scatter of abilities, with significantly higher abilities than mentally handicapped or normal children matched on mental age on tasks requiring discrimination of concrete visual-spatial relations, and significantly lower abilities than the other groups of children on tasks requiring abstraction (Maltz, 1981).

The autistic syndrome is marked by a number of behaviors that may include the following:

(a) lack of attachment behavior;
(b) lack of eye-contact;
(c) lack of verbal language, or deviant language marked by such abnormalities as pronomial reversal, and immediate or delayed echolalia;
(d) stimulus overselectivity;
(e) stereotyped and repetitive behaviors such as spinning or smelling objects;
(f) lack of symbolic play;
(g) self-stimulation (e.g., flapping hands, twirling) or self-abuse;
(h) a marked lack of awareness of the existence or feelings of others;
(i) no, or impaired imitation;
(j) insistence on sameness in the environment and marked distress in trivial changes in the environment.

(American Psychiatric Association, 1987; Rutter, 1978)

All of these behaviors have important instructional implications which are discussed later in the chapter.

Although most researchers believe that a basic cognitive deficit underlies autism and is fundamental to the disorder (Rutter, 1983), some investigators have suggested that the fundamental biological deficit has affective roots rather than cognitive ones, and recent research (Hobson, 1986a, 1986b, 1989, 1990) has provided interesting findings supporting this view. Details regarding the affective/cognitive debate are found elsewhere (e.g., Baron-Cohen, 1988, 1989e; Boucher, 1989; Fein, Pennington, Markowitz, Braverman, & Waterhouse, 1986; Frith, 1989a; Hobson, 1991; Leslie & Frith, 1990), and will not be discussed here.

Theory of Mind, its Precursers and Development in Autistic and Nonautistic Children

According to Leslie (1988), the development of a theory of mind depends on a specific innate mechanism that accounts for the emergence of pretend play in the second year of life (Leslie, 1987, 1988). Indeed, pretense is considered to be an early manifestation of, and precurser to the development of theory of mind (Leslie, 1987). It is in the second year of life that children begin to understand that one object can represent another object. For example, at this age, a child may see a banana, recognize it as a banana, pick it up, hold it to the side of his/her head and begin to simulate talking, thus pretending that it is a telephone. Once children are able to do this, they are able to "decouple", that is, they are able to separate their understanding that a banana truly *is* a banana in the "real" world but that it can represent something else when one wishes to engage in pretense (Leslie, 1987).

When children acquire the ability to pretend themselves, they simultaneously acquire the ability to understand pretense in others. Two year old children are able to understand that when others act as though a banana is a telephone, they are indeed pretending and actually *know* that in reality the banana is *really* a banana (Leslie, 1988). Thus, in a primitive form, children of this age *are* able to understand alternative realities. Despite this ability, they appear to be unable to understand instances of false-belief, situations where another person is led to believe something that is incorrect. Leslie (1988) attributes this ability difference to the situation itself. In the case of pretense, the alternative "reality" is invented, it is made up, whereas in the case of false-belief, there is a right and a wrong answer. The other person's belief has to be worked out on the basis of what that person did and did not see of the situation that led to the false-belief, and what that person will *do* on the basis of that belief. This information, Leslie (1988) notes, is not given to the children but must be calculated by them. Thus, the false-belief task is considerably more difficult for very young children to understand than pretense.

As will be seen, the majority of autistic children having mental ages (MAs) higher than that of the age when theory of mind emerges in normal children, appear to have a theory of mind neither in its primitive form (pretense) nor in its more advanced form seen in understanding of false-belief. Based on these findings Leslie (1987) suggested that autistic children are "decoupling impaired" in the latter domain. That is, they are unable to separate what actually *is*, from another person's *belief* of what a situation is.

Another aspect of development considered to be a precursor in the development of theory of mind, is perceptual perspective taking. Whereas theory of mind may be considered *conceptual* perspective taking, *perceptual* perspective taking "refers to an inference a child makes regarding another person's visual, auditory, or other perceptual experience" (Marvin, Greenberg, & Mossler, 1976, p. 511). Flavell, Everett, Croft, and Flavell (1981) postulated a developmental progression in perceptual perspective-taking, with children progressing from Level 1 to Level 2. Children at Level 1 are able to recognize that another person might be able to see an object that they cannot see or vice versa. However, children at Level 2 are also able to understand that an object that can be seen by both parties may appear differently to them depending on their relative locations, the location of the object, the object's distance, etc.

Research has shown that autistic children are successful in such perceptual role-taking tasks as a hide-and-seek task in which a miniature figure was said to be wishing to hide from two other figures who were the "seekers", and a task in which the subject was requested to turn a doll so that the doll could "see" the subject or the experimenter (Baron-Cohen, 1989c; Hobson, 1984). This result suggested that autistic children are able to recognize that "others see things differently" (Hobson, 1984, p. 102). Baron-Cohen et al. (1985) concluded that perceptual perspective taking tasks can be solved solely through the use of visuo-spatial skills, and mental rotation of representations, rather than belief attribution (Baron-Cohen, 1988, 1989c). Maltz (1981) obtained similar results to Baron-Cohen (1989c) and Hobson (1984) in a study that compared the cognitive deficits between autistic, mentally retarded, and normal children on the Arthur Adaptation of the Leiter International Performance Scales. Specifically, Maltz (1981) found that autistic children performed significantly better than either the mentally retarded or the normal subjects on tasks requiring concrete discrimination skills (e.g., counting concealed cubes, identical matching tasks, and matching number concepts or ordering blocks according to number). However, the mentally retarded and normal subjects performed significantly better than the autistic subjects on items requiring the use of abstraction and a comprehension of the relationships underlying stimulus configurations (e.g., matching according to family membership, sequencing blocks according to relationship of few to many, or part to whole).

Hobson (1984) observed that perceptual perspective taking tasks are the least "social" of perspective taking tasks. Perceptual perspective taking tasks may involve "seeing" another person's point of view, but it does not involve the more "personal" sensitivity required in understanding another person's thoughts, feelings, intentions or beliefs, an ability that *is* needed in conceptual perspective taking tasks.

Baron-Cohen (1989c) also investigated autistic children's ability to produce and comprehend protoimperative pointing (use of pointing to use another person to obtain an object), and protodeclarative pointing (use of pointing in order to comment or remark on the world to another person). In normal children, these joint-attentional gestures emerge in the second year of life, and appear to be another precursor to the development of a theory of mind (Baron-Cohen, 1989c). Baron-Cohen (1989c) found that autistic children were able to comprehend protoimperatives as capably as normal and Down syndrome children matched on mental age were able to comprehend them. However, the autistic children were significantly worse than the control groups in the comprehension of protodeclaratives. The autistic children were also found to be severely impaired in their production of protodeclarative pointing compared to the control groups, but not significantly different in the production of protoimperative pointing. Protodeclarative pointing has considerable significance in the development of a number of psychological abilities that include joint-attention behaviors, dialogue and symbol use, all of which are impaired in autism (Baron-Cohen, 1989c).

Details of experiments investigating perceptual perspective taking and protoimperative/protodeclarative pointing in autistic children can be found elsewhere (Baron-Cohen, 1989c; Hobson, 1984; Leslie & Frith, 1988). The point to be emphasized here is that autistic children appear to be impaired in the social aspects of these tasks, a finding that supports the theory of mind deficit hypothesis.

Symbolic Play

Since the existence of symbolic play is fundamental to the development of representational thought (Piaget, 1962; Riguet, Taylor, Benaroya, & Klein, 1981) and hence to the concept of theory of mind (Leslie, 1987, 1988), it will be discussed at some length.

Baron-Cohen (1987) noted the distinction that must be made between symbols and signs in determining the existence of symbolic play in autistic children. According to Baron-Cohen, a sign is a first order representation, or a primary representation of an object (e.g., a toy car). In other words, a sign indicates "the existence—past, present or future—of a thing, event, or condition" (Langer, in Baron-Cohen, 1987, p. 146). Symbols, on the other hand, are representations of concepts, or second-order representations (e.g., a representation of the *concept* of a toy car, rather than a representation of the toy car itself). In short, they are meta-representations. Using these definitions for sign and symbol, research has shown that autistic children are capable of representing the physical world, of making and using primary representations, and hence are capable of functional play (i.e., using objects "appropriately" according to their intended function) (Baron-Cohen, 1987; Lewis & Boucher, 1988; Sigman & Mundy, 1987; Wing, Gould, Yeates, & Brierley, 1977), although the quality of the functional play is poorer than that of normal or mentally retarded children. However, Baron-Cohen (1987) posited that autistic children demonstrate impairments in their ability to produce second-order representations (e.g., the ability to distinguish one's own belief from that of another person's *different belief* is assumed to be a skill requiring second order representations). "Thus, if a symbol is defined as a second-order representation, then the evidence suggests that autistic children *do not have the capacity to produce symbols*" (Baron-Cohen, 1987, p. 146).

Just as signs must be differentiated from symbols, Baron-Cohen (1987) and others (Leslie, 1987; Mundy, Sigman, Ungerer, & Sherman, 1987; Sigman & Ungerer, 1984) believe that symbolic play must be differentiated from functional or "reality" play (e.g., using a toy comb to comb a doll's hair). These authors believe that symbolic play occurs only when a child uses an object as if it were another object (e.g., "this banana is a telephone"), when a child attributes properties to an object which it does not have (e.g., "this doll's face is dirty"), or when a child refers to absent objects as if they were present (e.g., "this [empty] cup is full of tea"). A child demonstrating functional play behavior on the other hand, is merely representing an understanding of the physical world, not an understanding of the abstract world and the transformational quality of objects that is implicit in symbolic play (Baron-Cohen, 1987).

Most studies have found a significant paucity of symbolic play in the behavioral repertoire of autistic children (Riguet et al., 1981; Sigman & Mundy, 1987; Sigman & Ungerer, 1984; Wing et al., 1977). Autistic children tend to demonstrate ritualistic, stereotyped and manipulative play (such as spinning the wheels of toy cars, or repetitively loading and unloading toy trucks), far more often than do normal or mentally retarded children of equivalent mental ages (Wing et al., 1977), and tend to engage in imitative rather than symbolic play (Riguet et al., 1981).

Baron-Cohen (1987) matched autistic and Down syndrome children for MA and included a sample of nonhandicapped children whose chronological age (CA) matched the MAs of the other groups (mean CA = 4.1). The incidence of four categories of play was measured in spontaneous play situations. The four categories of play were sensorimotor play (banging, waving, sucking or sniffing objects with no attention paid to their "function"), ordering play (child imposes some pattern onto objects such as lining them up but with no regard for their "function"), functional play (using objects "appropriately" according to their intended function), and pretend play (child uses object as if it were another object, attributes properties to an object which it does not have, or refers to absent objects as if present).

Baron-Cohen (1987) found a significant difference between the autistic and the control groups in the pretend play category only. Significantly fewer autistic children produced any spontaneous pretend play relative to the control groups, although the incidence of functional play was not significantly different between the three groups. This finding lends support to Sigman and Ungerer's (1984) division of representational thought into two systems (one reflecting the development of sensorimotor skills, and the other reflecting the capacity to translate experience into language and play symbols that can be manipulated independently). Only one of these systems, the ability to form and manipulate symbols, is impaired in autistic children (Baron-Cohen, 1987, 1989b). This finding is confirmed by Hammes and Langdell (1981), and supports the theory of mind hypothesis. Since the incidence of pretend play in the Down syndrome group was not significantly different from the normal group, Baron-Cohen concluded that the lack of symbolic play is an autism-specific deficit.

As mentioned above, most studies, although not using Baron-Cohen's (1987) precise definition, have also found symbolic play impairments in autistic children (e.g., Wing et al., 1977). Lewis and Boucher (1988), however, arrived at different conclusions in their research. Using toys and "junk" objects, Lewis and Boucher measured spontaneous, elicited and instructed functional and symbolic play in autistic and normal children and children with moderate learning difficulties. They found that during the spontaneous play condition, autistic children's functional and symbolic play (assessed together), was as complex and varied as that of the control subjects, although the autistic group spent less time playing functionally and played with a greater number of toys per time unit than did the control groups. The autistic subjects were also able to carry out instructions for symbolic play. These results led Lewis and Boucher (1988; Boucher & Lewis, 1990) to the conclusion that although underused in spontaneous play, the *potential* for symbolic and functional play remains unimpaired in autistic children.

Other investigators (Riguet et al., 1981) have felt it unnecessary to differentiate between functional and symbolic play. Riguet et al. (1981) assessed autistic children's capacity for symbolic play, and attempted to determine the effectiveness of modeled symbolic play in eliciting higher level play. Each child was presented with animate toys (e.g., a plush monkey) and realistic and substitute accessories (e.g., plastic spoon and popsicle stick respectively). The children were given two free-play periods of four minutes at the beginning and end of each of two play sessions. These sessions were interspersed with two structured play periods in which the experimenter modelled symbolic play by having the toy (i.e., plush monkey) use the substitute accessories in a meaningful way. Riguet et al. (1981) found that the children were able to imitate modelled symbolic play, and concluded that autistic children's symbolic play can be improved. However, what was actually modelled and imitated appeared to be largely functional play (e.g., toy monkey made to hold plastic spoon or popsicle stick in its paw and bring it to its mouth tilting it as if sipping liquid from it). Whether imitation can be considered symbolic play, and whether the ability to imitate denotes an ability to improve symbolic play is open to question (Baron-Cohen, 1990b). Thus, lack of precision in the definition of symbolic play in many studies has led to a lack of precision in the interpretation of findings (Baron-Cohen, 1987).

Lewis and Boucher (1988) contended that methodological problems such as subject selection and use of play measures in Baron-Cohen's (1987) study accounted for the difference in findings between their study and his study. They noted that Baron-Cohen's study, among others in the area, did not use a measure of the duration of different kinds of play, the omission of which may have led to an overestimate of the normality of autistic children's functional play. Nor was there any measure of play quality in Baron-Cohen's study apart from distinguishing between different kinds of play which may have obscured the existence of impaired functional play.

Another explanation for the differences in results may lie in the choice of play materials in each of the studies. Baron-Cohen (1987) primarily used objects that were "realistic," thus, possibly limiting the type of play that was likely to occur. The children may have felt it was unnecessary to go beyond what the objects actually represented and create pretend scenarios. On the other hand, Lewis and Boucher (1988) used conventional toy materials as well as "junk" materials such as paper clips. Hence, more creative play may have been elicited since some of the materials did not possess inherent "play" characteristics.

The nonequivalence between the autistic samples in the two studies may also partially explain the contradictory results. Whereas Baron-Cohen's (1987) autistic sample was selected on the basis of verbal (vocabulary comprehension) and nonverbal (Leiter International Performance Scale) measures, resulting in a mean verbal MA of 2.5, and a mean non-verbal MA of 4.9, Lewis and Boucher's (1988) sample was selected on the basis of the same vocabulary comprehension measure and a measure of complex language. The resulting autistic sample for this study had a mean verbal MA of 5.9 and mean expressive language MAs of 5.5 and 4.3 in Information and Grammar respectively. Thus, it is possible that the two samples differed too greatly to be comparable. Lewis and Boucher's (1988) autistic sample may have been higher functioning and thus, may have had more highly developed symbolic play skills.

To summarize, the development of symbolic play is thought to be an important precursor to the development of theory of mind (Leslie, 1987, 1988). The simplest form of symbolic play, functional play, concerns the representation of objects (i.e., primary representations or signs). However, in the more complex forms of symbolic play (object substitution, attribution of properties, and creation of non-existent objects), the representations are far more abstract. The representations are second-order (Baron-Cohen, 1987), similar to the meta-representations required for a fully functional theory of mind. Both Sigman and Mundy (1987) and Rutter (1987) posited that these representational deficits may form the basis of autism's cognitive deficits.

Theory of Mind Research with Autistic Children

Research on theory of mind with normal and autistic children has included studies of children three to twelve years old, and has dealt with numerous aspects of the phenomenon including first- and second-order false-belief attribution (Baron-Cohen, 1989d; Baron-Cohen et al., 1985; Gopnik & Astington, 1988; Hogrefe, Wimmer, & Perner, 1986; Perner et al., 1989; Perner & Wimmer, 1985; Perner, Leekam, & Wimmer, 1987; Wimmer & Perner, 1983), appearance–reality distinction (Baron-Cohen, 1989a; Flavell et al., 1983; Flavell et al., 1986; Gopnik & Astington, 1988), representational change (Gopnik & Astington, 1988), knowledge of the existence of other mental states (Baron-Cohen, 1989c; Baron-Cohen et al., 1986; Leslie & Frith, 1988; Marvin et al., 1976; Perner et al., 1989; Ruffman & Olson, 1989), and causes of emotion (Baron-Cohen, 1991a). It appears that these abilities, all of which emerge at about the same age for normal children, are different, but developmentally related acquisitions that are expressions of some underlying new ability (Astington & Gopnik, 1988; Flavell, 1988). Theory of mind studies conducted with autistic children will now be discussed with reference to their contribution to an understanding of the nature of the theory of mind deficit. Appendix A provides a summary of these studies with specific demographic data on the subjects used in the various studies.

False-Belief Studies

In attempting to determine the presence of a theory of mind in autistic children, Baron-Cohen et al. (1985) examined these children's ability to solve a false-belief problem. The researchers used three groups of children in the study: autistic (mean CA = 11.11), Down syndrome (mean CA = 10.11) and normal (mean CA = 4.5), with the Down syndrome and normal subjects acting as control groups. The autistic group had a mean IQ of 82 and the Down syndrome group had a mean IQ of 64. Relatively high functioning autistic subjects were chosen to enable the researchers to test the hypothesis of an autism-specific deficit.

In this experiment, the children were shown two doll protagonists, "Sally" and "Anne." In the scenario, Sally hid her marble in a basket and then went for a walk. While she was away, Anne transferred the marble to a box. The subjects were asked to determine where Sally would look for the marble when she returned (a Belief question). If they were able to understand Sally's perspective, then they should have pointed to the basket and not to where the marble *really* was. The subjects were asked a number of control questions in the course of the experiment: a naming question, to ensure that they understood which doll was which, a reality question, to ensure they had knowledge of the current real location of the marble, as well as a memory question, to ensure that the children possessed an accurate memory of the marble's original location. The experiment was repeated using new locations for the marble, so that there were three different potential locations to which the children could point (basket, box, or experimenter's pocket).

All the subjects in the study passed the control questions on both trials. Whereas 85 % of the normal children (23 out of 27) and 86% (12 out of 14) of the Down syndrome children passed the belief question on both trials, 80% (16 out of 20) of the autistic children failed the belief question on both trials ($p < .001$). All of the autistic children who failed the belief question pointed to where the marble actually was. Baron-Cohen et al. (1985) concluded that the autistic children did not appreciate the difference between their own and the doll's knowledge, and that they were unable to represent mental states. In short, the autistic group was unable to employ a theory of mind to solve the task, an inability that appeared to be an autism-specific deficit, given the success of the more severely retarded Down syndrome children.

Two possible alternative explanations may account for the autistic children's performance in this study. First, it is possible that the pretend nature of the task interfered with the autistic group's ability to correctly answer the Belief question. The possibility has been argued that the use of dolls may have made the task more difficult for the autistic children since they are known to have symbolic play deficits (de Gelder, 1987). Hence, the autistic subjects may have been unable to suspend their knowledge of what actually *was* (a decoupling impairment), to pretend that a doll could have any perspective at all. As will be seen in the description of a subsequent experiment, however, it is unlikely that the use of dolls played a role in the autistic children's failure at this task.

A second explanation concerns the salience of the stimulus, and recency effects. In this study, the last location of the marble was always the wrong answer and was the location consistently chosen by the autistic children. While it is possible that this did not make a difference to the Down syndrome and normal children, it may have interfered with the task for the autistic children. The autistic children may have focused only on the most recent location of the marble, or only on where the marble *actually* was, since this was the most salient characteristic of the situation. Cognitive aspects of the situation were more likely to be ignored simply because they were less overt, observable aspects than the actual location of the marble (Wellman, 1985). Hence, the results of this study may only suggest that autistic children *do not* employ a theory of mind, not that they *cannot* employ one. The autistic children's performance, then, may actually reflect stimulus overselectivity (American Psychiatric Association, 1987; Hermelin & O'Conner, 1970) rather than a theory of mind deficit.

In a second study, Baron-Cohen et al. (1986) followed up the same children to obtain additional evidence regarding the autism-specific nature of the theory of mind deficit, and to compare autistic children's understanding of physical and social events directly. In this study, Baron-Cohen et al. (1986) used a picture sequencing task in which the children were to sequence correctly three of four pictures in each story and narrate the stories. The first picture in the sequence was presented to the children to reduce their possible ambiguity. Three types of picture stories were utilized. The first was "mechanical" in which physical-causal relations were depicted (e.g., Picture 1-Rock on hilltop, 2-Rock topples, 3-Rolls down hill, 4-Knocks tree over). The second type depicted "behavioral" sequences (e.g., Picture 1-Boy turns on tap, 2-Stands under, 3-Soaps himself, 4-Dries himself), while the third type of sequence depicted "Intentional" stories (e.g., Picture 1-Boy puts chocolate in box, 2-Goes out to play, 3-Mum eats chocolate, 4-Boy sees chocolate gone). To rule out the possibility that difficulties in understanding the Intentional stories were due to problems with the social content of the scenario, Baron-Cohen et al. (1986) included sequences in the mechanical category that involved only objects as well as ones that involved persons and objects. In the behavioral category, sequences depicting events involving only one person were compared with those involving two people.

Baron-Cohen et al. (1986) found that performance on the mechanical and behavioral subconditions comparing object versus person, or single person versus person-to-person, was highly similar for all groups, thus ruling out the possible confounding effects of the "social" nature of the tasks. For all subsequent comparisons, then, these subconditions were collapsed. The results indicated that the autistic children's performance was significantly superior ($p < .001$) to both the normal and Down syndrome groups in the mechanical condition, presumably attributable to the autistic children's higher MA. In the behavioral condition, the autistic and normal groups performed virtually identically, with the Down group performing significantly worse. In the Intentional condition, however, the autistic group performed significantly poorer than the Down group ($p < .05$), despite the higher mean nonverbal and verbal MA of the autistic group. Baron-Cohen et al. (1986)

believed that these results confirmed and extended the findings of their previous study (Baron-Cohen et al., 1985). Because the autistic group was able to correctly sequence mechanical stories involving a person, and behavioral stories involving interactions between people, Baron-Cohen et al. (1986) contended that autistic children demonstrated a *specific* social deficit in employing a theory of mind rather than a general social deficit. In the story narrations, the autistic children (with two exceptions) gave purely descriptive renditions, while children in the other groups gave either implicit or explicit mental state explanations.

The question of the specificity of the deficit in autistic children was further explored by Baron-Cohen in a more recent study (1991b). In this study Baron-Cohen used a number of tasks to investigate whether or not autistic children's deficit in attributing mental states to others extended to areas of social cognition which did not involve a theory of mind, namely relationship recognition (categorizing pictures depicting four categories of relationships—mother–child, father–child, peer, and husband–wife), interpersonal reciprocity (distinguishing own left–right from other left–right when sitting opposite another person, and maintaining a ball-rolling game), and understanding the animate–inanimate distinction (categorizing pictures of animals, plants, people, domestic objects, mobile objects, and toy creatures as either alive [animate] or not alive [inanimate]).

The results of the study indicated that relative to normal and mentally handicapped control groups, autistic children were unimpaired in all three domains of social cognition. On the test of social relationship recognition, the autistic group scored significantly higher than the mentally handicapped subjects, but did not differ significantly from the normal group. On the reciprocity tasks the three groups did not differ from one another statistically. On the final task, the autistic group scored well above chance in sorting both the pictures of animate and inanimate objects, while the mentally handicapped and normal subjects, although well above chance on their inanimate scores, were only just above chance on their animate scores. Baron-Cohen (1991b) concluded that since these social cognition tasks did not require the attribution of mental states, the theory of mind deficit in autistic children is highly specific to tasks that do require such an attribution.

The theory of mind studies discussed thus far have demonstrated that autistic children seem to be impaired in understanding false-belief, a major aspect of theory of mind. Leslie and Frith (1988) sought to confirm the results of the 1985 Baron-Cohen et al. study by replicating the results with different autistic children, and by using real people, rather than dolls, to act out the scenarios. Similar results would confirm the belief that the use of dolls in the Baron-Cohen et al. (1985) study was not a confounding factor. Leslie and Frith (1988) also sought to extend the previous findings by examining autistic children's ability to understand the concepts of seeing and knowing, as well as believing. To test the autistic children's ability to remember the location of an object, an ability that plays an important role in the mental state tasks, a memory task was also included.

In Leslie and Frith's (1988) study, children with specific language impairments (SLI) were matched on verbal MA with a group of autistic children (*n* = 12 and *n* = 18 in the respective groups) and four tasks were presented: line of sight, memory for position, limited knowledge and false-belief. The line of sight task consisted of a plastic board that was placed on the table perpendicular to the child, and between the child and the experimenter. A doll was placed on one side of the board (always visible to the child on each trial) and a counter was introduced to one or the other side of the board. The child was then asked if the doll could see the counter. On subsequent trials both the doll's, and the counter's position was varied. (If any children had difficulty in accepting that the doll could "see", a co-experimenter took the place of the doll).

The memory task tested the children's ability to remember where a marble's original location was after it had been removed from the location and handled for a few seconds. All of the autistic children passed both of these tasks without difficulty.

The limited knowledge task involved having one experimenter hide a counter in the presence of another experimenter and the child. The second experimenter then left the room, and the first experimenter produced another counter and had the child hide it in a different location. Two control questions ("Can you put it [the counter] somewhere different?" and "Where did [experimenter 2] see me hide a counter?") were asked to ensure that the child was able to follow the procedure. The first experimenter then asked the child if the second experimenter would know where the second counter was hidden (knowing question) and where the second experimenter would look for a counter upon return (prediction question). While the autistic children were able to correctly answer the control questions, 10 of the 18 children failed the criterion of a correct response to both the knowing and the prediction questions.

The final task in this study was the false-belief task. In this task, experimenter 1 had experimenter 2 hide a 1£ coin in view of the child and then leave the room. Experimenter 1 had the child indicate the location of the coin. The experimenter then removed the coin and hid it in a different location. The child was then asked a series of control questions, a knowing question, a prediction question, and a think question ("Where does [experimenter 2] think the coin is?"). Again, the children in both groups were able to answer the control questions correctly, and the 12 SLI children passed the prediction and knowing questions, with one child failing the thinking question. In the autistic group, however, five children passed and 13 failed the prediction question. Passing or failing on the thinking question was completely consistent with passing or failing the prediction question for the autistic group. Thus, the difference between the SLI and autistic groups was significant ($p < .001$).

Leslie and Frith (1988) concluded that their results provided further support to the idea that a majority of autistic children have difficulty with mental states, and that they experience a gross delay in the development of a theory of mind. Although the methodological refinements of this study, including the use of real people, and constant checks to ensure that the children understood the procedure, strengthened the conclusions, a few uncontrolled rival hypotheses remain. For example, the memory task included in this study did not rule out the stimulus salience hypothesis stated earlier regarding the reason for the autistic children's failure on the false-belief task. In the memory task there was only one location in which the marble had been hidden. Despite the time delay, the correct answer was always the marble's last location, whereas in the false-belief task, the marble's last location was always an incorrect answer. Autistic children succeeded on the memory task and failed on the false-belief task while always selecting the *last* hiding place of the marble. Although this stimulus salience hypothesis may be unlikely for the false-belief task in this study (given the correctly answered control question in which the child was asked about where the second experimenter had put the coin in the first place), its possibility should be entertained and controlled for, as it has in similar studies conducted with normal children (Flavell et al., 1983).

Perner et al. (1989) sought to replicate Baron-Cohen et al.'s (1985, 1986) findings that autistic children's theory of mind is impaired, and improve the experimental paradigm at the same time. Perner et al. (1989) argued that rather than failing to understand that false-belief arises from an unexpected change in the world, the autistic children's failure in Baron-Cohen et al.'s (1985, 1986) studies "may reflect nothing more than a difference in common-sense assumptions about what people normally expect to happen" (p. 690). Hence, Perner et al. (1989) employed a "deceptive appearance" paradigm in which the subjects first experienced how they themselves were misled about the contents of a "Smarties" box by its appearance. Perner et al. (1989) also investigated autistic children's ability to attribute mental states other than false-beliefs (i.e., knowledge and ignorance), the relationship between these various aspects of theory of mind, and the pragmatic skill of making a communicative adjustment to the knowledge of the listener.

Subjects for this experiment included 26 children diagnosed as autistic (mean verbal MA = 6.2), and 12 children with specific language impairments (SLI) with a mean verbal MA of 6.9. The autistic children were tested on four tasks: communication test: "Boxes", false-belief, communication test: "Bee", and a knowledge-formation task, while the SLI children were tested on two false-belief tasks only, the deceptive appearance task used with the autistic children and an adaptation of the paradigm used by Baron-Cohen et al. (1985).

In the boxes communication test, the main experimenter showed the subject two small boxes of differing size and informed the subject that they would hide some things from a co-operating experimenter who was also present. Two task conditions were used in the test: the Total Ignorance condition, and the Partial Ignorance condition. In the Total Ignorance condition, the co-operating experimenter remained unaware of the contents of either box, while in the Partial Ignorance condition, the co-operating experimenter was eventually able to open one of the boxes. Depending on the task condition, the co-operating experimenter either left the room at this point (Total Ignorance condition), or remained in the room (Partial Ignorance condition), as the main experimenter produced a wax apple (considered a salient, interesting item by the experimenters), asked the subject what it was, and hid it in one of the boxes. If the task condition was the Partial Ignorance condition, the co-operating experimenter left the room at this point. A piece of crumpled paper (considered a dull item) was then produced, the subject was asked what it was, and the paper was then hidden in the other box. The co-operating experimenter returned and tried (unsuccessfully) to open the box containing the paper. The co-operating experimenter then tried to open the second box. In the Total Ignorance condition, the co-operating experimenter was to remain unaware of the contents of either box, and thus was unsuccessful in opening the second box as well. In the Partial Ignorance condition, however, the co-operating experimenter was able to open the second box (containing the wax apple), and announced its contents. The box was then closed and the co-operating experimenter asked the subject, "What's in there?", without indicating which box was the focus of the question. The subject's answer to this question determined whether or not the subject was able to take account of the co-operating experimenter's limited knowledge. If the subject mentioned the contents of only one box, memory for the contents of the second box was checked by asking "What is in the other box?".

In the bee communication test, the experimenters used a mechanical bee that could flap its wings (considered interesting and salient) and nod its head (considered dull) when turned on. In this task, the subject was required to tell a previously absent co-operating experimenter what a toy bee could do upon his or her return. This task was similar to the boxes task, in that the co-operating experimenter was in either a Partial Ignorance condition (i.e., left the room after the demonstration of the bee "flying" by flapping its wings) or Total Ignorance condition (i.e., left the room before any of the bee's actions were demonstrated).

The results of the communication tasks indicated that only 12.5% of the autistic children reliably took into account the listener's knowledge in shaping a message, even though many of the subjects were capable of working out what that knowledge might be. A supplementary experiment, in which the Partial Ignorance condition was replicated using different objects, involved retesting 20 autistic subjects' capability of determining what information the co-operating experimenter required. Whereas on the initial test, only three subjects correctly identified the dull item (the crumpled paper) as the item unknown to the co-operating experimenter, seven subjects were able to correctly identify the dull item upon retesting, thus demonstrating that more autistic children are *able* to take account of a listener's knowledge than actually utilize their ability. Perner et al. (1989) concluded that this finding may indicate that even where autistic children have a certain level of facility in understanding mental states, they may not fully utilize it in communication situations.

In the false-belief task, the main experimenter showed the subject a Smarties box and asked, "What's in here?" All subjects answered, "Smarties". The experimenter then opened the box to reveal that a pencil, not Smarties, was in the box. The experimenter replaced the pencil in the box and asked the subjects what was in the box, and what the subjects had answered the first time they were asked. Each subject was then asked what the next subject would believe is in the box upon seeing it. Similar to the Baron-Cohen et al. (1985) study, only 4 of the 26 autistic subjects were able to make correct predictions regarding another child's belief in the false-belief task. All but one of the SLI subjects were able to make correct predictions on both false-belief tasks.

For the knowledge-formation task, the main experimenter hid an object in a cup. In the Other Ignorant condition (Other), the experimenter permitted the subject to see what was in the cup without allowing the co-operating experimenter to see the cup's contents. In the Subject Ignorant condition (Self), however, the experimenter allowed the co-operating experimenter to see the cup's contents, but not the subject. In both conditions, the subject was then asked a series of questions relating to knowledge attribution (e.g., "Does [the co-operating experimenter] know which thing I put into the cup?", "Why does [co-operating experimenter] not know that?", "Did I let [co-operating experimenter] look into the cup?").

Most of the autistic children were able to evaluate visual access correctly (N = 17 for Other and 16 for Self). That is, most of the subjects were able to identify who had been allowed to see the object. However, the subjects were much less able to make correct knowledge attributions (N = 10 for Other and 13 for Self). All 10 subjects who were able to make correct knowledge attributions were also successful in judging the other person's visual access, while the 6 children who failed in judging the other person's visual access, also failed in knowledge attribution, resulting in a significant positive correlation (*phi* =

.52, Fisher's test: $p < .05$). The remaining 7 children who were able to make correct judgements of Other's visual access, failed to attribute knowledge correctly. No subjects showed the opposite pattern. The contingency between correct knowledge attributions to Other and Self was almost perfect (*phi* = .91, Fisher's test: $p < .001$). Since eight subjects gave adequate justifications for Other and Self on both tasks, Perner et al. (1989) concluded that about 35% of the children tested were able to make a clear connection between visual access and knowledge. There was also a strong contingency between belief and knowledge attribution. The four subjects who made correct false-belief attributions were also able to attribute knowledge on the basis of visual access and justify their attribution. All 12 children who failed on knowledge attribution also failed on belief attribution. The resulting correlation was positive and statistically significant (*phi* = .61, Fisher's test: $p < .05$).

Perner et al. (1989) concluded that their study supported the claim that able (as defined by mental age) autistic children are severely impaired in their theory of mind, and that this is an autism-specific deficit. The investigators concluded, as well, that certain aspects of theory of mind are more easily developed than are others. Specifically, more autistic subjects in this study were able to understand "knowing" in terms of visual access than were able to attribute false-belief, and all four children who understood false-belief demonstrated understanding of the knowledge–ignorance distinction. However, another four children demonstrated understanding of the knowing/not-knowing distinction without being able to understand false-belief. Thus, it appears that autistic children found the knowing/not knowing distinction slightly easier to understand than the false-belief task.

The final false-belief study to be considered in detail is another one by Baron-Cohen (1989d), in which the specific delay hypothesis regarding autistic children's theory of mind acquisition was extended. In this study Baron-Cohen investigated the ability of autistic, Down syndrome, and normal children, who were capable of first-order belief attributions (i.e., the 20% of autistic children who were successful in the 1985 Baron-Cohen et al. study, to make second-order belief attributions (i.e., of the form "Mary thinks John thinks the ice-cream van is in the park"). The inclusion criterion for the autistic children was the ability to pass the simpler first-order belief attribution task. The autistic group was specifically chosen to have higher verbal and nonverbal MAs than those of the control groups in order to test the hypothesis that there was a *specific* deficit in the autistic group's ability to make second-order belief attributions, a deficit independent of a *general* developmental delay.

In this study, a toy village was used, in which toy characters and props (John, Mary, the ice-cream man, and his van) were moved about the village according to a story. The critical aspect was that by the end of the story Mary possessed incomplete information as to the "true" location of the ice-cream van, and thus should have thought that John believed something different than what was actually the case (i.e., that Mary thought that John thought that the ice-cream van was still in the park, [though in actuality, the van was now parked at the church]). The subjects were asked a number of prompt, or control, questions throughout the task to ensure they were following and understanding the story, and were asked four questions at the story's conclusion: a belief question (i.e., Where does Mary think John has gone to buy an ice-cream?), a justification question, in which the children had to justify their belief (i.e., Why?), a reality question (i.e., Where did John really go to buy his ice-cream?), and a memory question (i.e., Where was the ice-cream man in the beginning?). The purpose of the justification, reality, and memory questions was threefold: (a) They acted as control questions to provide evidence of which level belief attribution the children were making, (b) the questions ensured that the children had knowledge of the real location of ice-cream van, and (c) the questions ensured that the children had an accurate memory of the van's previous location.

As in previous studies, all of the children (with the exception of one autistic child) passed all the prompt questions, as well as the reality and memory questions. While 90% of the normal group and 60% of the Down syndrome group passed the belief question, none of the autistic subjects did so ($p < .01$). All of the autistic subjects pointed to where the van really was (the church) rather than to where Mary thought John believed it was (the park). All of the subjects who passed the belief question were able to answer the justification question by attributing a second-order belief (i.e., "[Mary] thinks [John] thinks it's in the park."). Those subjects who failed the belief question, however, either gave first-order belief attribution answers to the justification question (e.g., "[John] knows the ice-cream man is at the church.") or zero-order answers (i.e., reporting where the van actually was).

Baron-Cohen (1989d) concluded that autistic children who have developed a theory of mind at a lower level, are specifically delayed in the acquisition of a more complex theory of mind. Baron-Cohen (1989d) believed that the paradigm of this experiment was a closer model of the complexities of "real" social interaction than were his earlier experiments and, thus, more conclusive in terms of the specific delay hypothesis.

Causes of Emotion

Baron-Cohen (1991a), tested autistic children's understanding of some of the causes of two basic emotions (happiness and sadness). Developmental literature suggests that three to four year old children are capable of understanding that emotions can be caused by situations and desires, and by age six, most children are capable of understanding beliefs as a cause of emotion (Baron-Cohen, 1991a). Given autistic children's previously demonstrated impairment in understanding of belief, Baron-Cohen (1991a) hypothesized that the autistic subjects in this study would be impaired in their ability to understand belief as a cause of emotion.

Three groups of subjects, autistic, mentally handicapped, and normal, were matched on verbal MA (means of 6.9, 6.5, and 5.3 respectively). The autistic group had a higher mean nonverbal MA than the other groups of subjects (8.5 as compared to 6.0 and 5.3), thus making it unlikely that any impairment would be due to a general developmental delay. Each subject was introduced to a doll and told a story in which two questions testing understanding of situations as causes of emotions (Situation Test), two questions testing understanding of desires as causes of emotions (Desire Test), and two questions testing understanding of beliefs as causes of emotions (Belief Test) were asked. The latter test was a false-belief task similar to that employed by Baron-Cohen et al. (1985), the difference being that the subject was asked how "Jane" will *feel* upon discovering that what she thinks is in a box is not in the box. Each test question was followed by a Justification Question (i.e., "Why [did Jane feel the way that she did?]"). A second trial with minor variations, was then administered to each subject. In this trial a different doll (Mary) had the opposite desires to Jane. Memory and naming questions were included in the Desire Test to control for the possibility that memory or recognition problems might account for failure on the other questions. A second Desire Test was administered to control for order effects.

All subjects passed the control questions and the Situation Test on both trials. The normal group performed better than did the clinical groups on the Desire Test (92.1% versus 57.4% and 59.4% of the autistic and mentally handicapped groups respectively), with the difference approaching significance (normal x mentally handicapped, Chi square = 3.58, 1 df, $p = 0.058$). The difference between the two clinical groups was not significant. On the Belief Test, however, the autistic subjects performed significantly poorer than the other two groups with only 17.6% of the autistic subjects passing, compared to 56.3% of the mentally handicapped group and 73.7% of the normal group. The difference between the mentally handicapped group and the normal group was not significant. Thus, the results indicated that the autistic and normal groups found the Desire Test easier than they did the Belief Test, while for the mentally handicapped group, there was no apparent difference in difficulty between the two tests.

Based on the results, Baron-Cohen (1991a) concluded that it is when emotion and cognition interact that autistic children have difficulty in understanding emotion, and that autistic children alone are impaired in their ability to form metarepresentations, thus supporting Leslie's (1987) theory. Whereas understanding of emotions based on an understanding of desires and situations do not necessarily require metarepresentation, understanding belief as a cause of emotion does require metarepresentation.

From a developmental perspective, it appears that understanding desire as a cause of emotion develops prior to the understanding of belief as a cause of emotion for autistic children. There was a significant difference in chronological age between the autistic children who passed the Desire Tests but failed the Belief Test (t = 5.61, df=7, $p < .001$), with the older children being more successful on the Belief Test.

Appearance–Reality Distinction Studies

Only one study to date has examined the existence of appearance–reality distinction in autistic children (Baron-Cohen, 1989a). Baron-Cohen (1989a) used groups of autistic, mentally handicapped and normal children (n = 17, 16, and 19, respectively), with nonverbal MAs of 8.48, 6.03, and 5.3 and verbal MAs of 6.91, 6.47, and 5.3 in the respective groups. The MAs of the subjects in the normal group were not measured, but were assumed to be very close to their chronological ages. This experiment provided the subjects with a "warm-up" procedure in which the children were sensitized to the fact that things can appear one way, but in reality, be something different. The subjects were also given a memory pretest, in which changes in an object's apparent color were produced by placing an orange filter over a white sheet of paper. The children were then asked, "When I take this (filter) away, will the paper look white or orange?". Passing of this test constituted the inclusion criteria.

The subjects were presented with four randomly ordered appearance–reality tasks, in which an object's color, size, material, and identity were separately manipulated. Each task contained two trials (each trial using a different object), and the subjects' response to a pair of Appearance and Reality questions were coded as either correct, "phenomenist" (e.g., saying a sponge [painted to look like a rock] looks like a rock and really is a rock), "realist" (e.g., saying the object looks like a sponge and really is a sponge), or "other". The Color task involved a bottle of milk being presented to the child and then covered with an orange filter. The child was asked what color the milk looked (Appearance question) and what color it really was (Reality question). For the Size task, each subject was presented with two British coins of different sizes (a penny and a 10p piece). A magnifying glass was held over the penny making it appear larger than the 10p, and the subject was asked the appearance–reality pair of questions ("Now how does the penny look? Smaller or bigger than the 10p?" and "What size is the penny really? Bigger or smaller?" respectively). In the

Material task, the experimenter placed a plastic, but realistic looking chocolate, in front of the subject and asked the subject what it was. All subjects said it was chocolate. The children were then allowed to manipulate the "chocolate" to discover its true property and asked what it really was. All children were able to identify that it was plastic. The children were then asked the pair of Appearance–Reality questions ("What does it look like?", and "What is it really?"). In the Identity task, a stone, realistic-looking egg was shown to the children and the children were asked "What is this?" Like the material task, the children were permitted to handle the "egg", discover its true identity, and report what it really was. The children were finally asked the Appearance and Reality questions ("What does it look like?" and "What is it really?"). Each task was repeated using a piece of white chalk, a £1 coin and a 50p coin, a plastic, realistic-looking hamburger, and a sponge that looked like a slice of bread for the color, size, material and identity tasks respectively.

Results of the study showed that 78.9% of the mentally handicapped group, 81.3% of the normal group and only 35.3% of the autistic group passed the appearance–reality tasks overall. A task by task analysis showed that the four tasks were not significantly different for the normal and mentally handicapped groups. For the autistic group, however, the color task was easier than the others which were of uniform difficulty. Overall, 75% of the autistic group's errors were phenomenist errors. That is, 75% of the autistic children reported that the objects really were what they appeared to be, or that the apparent color of an object under a filter was the true color of the object, compared to 29% and 54% of the normal and mentally handicapped groups respectively making phenomenist errors on the material task.

The results of this experiment replicated the findings of Gopnik and Astington (1988), Flavell et al. (1983), and Flavell et al. (1986), among others. That is, the appearance-reality distinction is quite well developed in most normal children by six years of age. The finding that the mentally handicapped group performed well also supported the findings of studies investigating other aspects of theory of mind using this population (Baron-Cohen, 1989d; Baron-Cohen et al., 1985, 1986). Of most importance perhaps, was the finding that autistic children were impaired in the appearance–reality aspect of theory of mind as well. This finding demonstrated that when perceptual information was contrary to their knowledge about the world, autistic children were unable to separate the appearance from the reality. The perceptual information apparently overrides other representations of an object, and is an autism-specific deficit. Baron-Cohen (1989a) suggested a number of other possible interpretations of the findings, including the possibility that nonautistic children interpreted the pragmatics of the Reality question differently, and that autistic children might have benefitted if more assistance in recognizing the point of the task was provided.

Appearance-Reality Research with Non-Autistic Children

To more clearly understand the appearance-reality distinction phenomenon, a few experiments that have investigated the development of appearance-reality distinction in normal children will be briefly considered. In three experiments (each one using different children), Flavell et al. (1983) investigated three to five year olds' ability to distinguish between, and correctly identify real, as opposed to apparent object properties (color, size, and shape), object identities (e.g., realistic looking rock that is really a sponge), object presence–absence (i.e., a large object placed in front of a small object, with child being asked if both objects were there), and action identities (i.e., Action A resembled Action B when viewed from a certain perspective). The experimenters found that, although unstable, even three year olds had some ability to make correct A–R distinctions, and that this ability increased with age, but even the older group of children did not reach ceiling performance on the tasks. Like Baron-Cohen (1989a), Flavell et al. (1983) also found that phenomenism errors predominated on A-R tasks involving object properties. However, on tasks in which object identities, object presence-absence, and action identities were in question, realism errors predominated.

Flavell et al. (1983) included a brief training/feedback procedure for those children who failed the final task in Experiment 1, and in Experiments 2 and 3 a pretraining procedure was included to familiarize the subjects with the tasks. In Experiment 1, the feedback procedure involved having the experimenter hold up the last prior object about which the subject had answered correctly (i.e., correctly answered the Appearance and Reality questions), and said, "Remember you said this *looked* to your eyes like a _____, but it was *really, really* a _____. You were right. It does look to your eyes like a ____, but it really, really is a _____, and all of the objects we have talked about are the same way" (p. 102). The experimenter then labelled the appearance and reality of the other two objects and retested the subject on the fourth object.

The pretraining procedure in Experiment 2 consisted of demonstrating the meaning of the words "looks like" and "really, really" as follows. The child was presented with a sheet of green plastic that covered a small square of white paper with serrated edges. The child looked at the paper through a magnifying glass and the experimenter said,

> When you look at this with your eyes right now, it looks like a piece of cloth that is green and big. It looks like a big, green piece of cloth. But it isn't really. It's really kind of little (magnifier removed), and it's really, really white (green plastic removed), and it's really, really a piece of paper (paper handed to child to feel). It's really a white piece of paper that is kind of little. But when you looked at it with

> your eyes before, it looked like a piece of cloth (paper replaced on black table surface) that was green (green plastic replaced on paper) and big (magnifier placed between child and paper). Sometimes things look like one thing when they're really something else. Right? (p. 105)

Pretraining in Experiment 3 was slightly more extensive than that of Experiment 2, in that it involved two pretraining tasks instead of just one. In this experiment, after the subject was shown a Charlie Brown puppet, the experimenter put a handkerchief with eyes and mouth made of felt over the puppet and said,

> When you look at this with your eyes right now, it looks like a ghost. It looks like a ghost to your eyes. But it really, really isn't. It's really, really Charlie Brown. Sometimes things look like one thing to your eyes when they are really, really something else (p. 110).

In the second pretraining task, the child was given a piece of heavily starched cloth to feel. The experimenter then took the cloth and held it about a metre away from the child and said,

> When you look at this with your eyes right now, it looks soft and easy to bend. It looks soft and bendy, like this. (Experimenter holds up and moves an unstarched piece of the same cloth.) But it really, really isn't. It's really, really stiff and hard. Sometimes things look like one thing to your eyes when they're really, really something else (p. 110).

All of these training procedures, however, had little effect on the children's performance. The subjects were still unable to correctly answer the Appearance and Reality questions. Flavell et al. (1983) concluded that the children were experiencing more than a mere misunderstanding of the task demands or the A–R terminology.

In a much earlier study, Braine and Shanks (1965) found that correct responding on Size appearance-reality distinction tasks of two groups of preschool children (mean CA 3.11 and 4.10 respectively) improved significantly when concrete reinforcement, in the form of "winning" and "losing" poker chips contingent on correct responding, was provided during a brief training period. The youngest group of children (mean CA of 3.1), however, showed no evidence of learning, despite the use of concrete reinforcement. Braine and Shanks (1965) concluded that the subjects could not have learned such a complex distinction in a short laboratory teaching session, but that they were already cognizant of the distinction. The increased information feedback that the concrete reinforcement provided merely helped to clarify the task requirements for the subjects.

In 1986, Flavell et al. conducted a series of seven experiments investigating the nature and development of the A–R distinction from three years of age to adulthood. The specific experiments to be discussed here all used subjects with a mean age of 3.7. In two of these experiments, Flavell et al. (1986) included appearance–reality tasks considered to be easier than the "standard" ones for tasks designed for the three year olds. For example, in addition to such standard A–R tasks as the sponge/rock one used in Flavell et al. (1983), they also used a task in which milk (a familiar liquid) was made to look red (a color it had never had) by putting it behind red filter paper. Flavell et al. (1986) found that even the use of easier tasks and the inclusion of specific pretraining (i.e., the Charlie Brown puppet task described earlier) failed to yield any improvement in the children's performance.

In a further investigation regarding the effectiveness of specific appearance–reality instruction, Flavell et al. (1986) used different objects to illustrate how they appear one color when they are behind a colored filter, and another color when they are not behind a filter. The child was asked repeatedly what color the objects were "really and truly," and what color the objects "looked" when they were behind the filters. The results indicated that only one child (out of 16) appeared to profit from the instruction upon post-testing.

Similar results were found when tasks involving disguise, sound and smell were used. In one experiment, Flavell et al. (1986) included A–R tasks in which a can, when turned over, sounded like a cow, a balloon that sounded like a horn when deflated, a cloth that smelled like an orange, a sponge that smelled like a lemon, and a disguise (a full pull-over-the-head mask of a familiar creature ["Big Bird"] donned by a second experimenter. Flavell et al. (1986) thought that these tasks might be easier for pre-schoolers than standard A–R tasks since different sense modalities were involved. For example, although the can *sounded* like a cow (auditory modality to exhibit appearance) it was *really* a can (visual modality to exhibit reality). Flavell et al. (1986) found that although performance on the Disguise, Sound, and Smell tasks was significantly better than performance on a standard Object Identity task, it was not significantly better than performance on a standard Color task, and was poor in absolute terms as well, considering the very specific pretraining that the children were given. Both phenomenism and realism error patterns were evident in the Disguise, Sound and Smell tasks. That is, while some subjects gave appearance answers to both questions (phenomenism errors), other subjects gave reality answers to both questions. Flavell et al., (1986) concluded that three year old children can have as much difficulty understanding auditory and olfactory appearance–reality distinctions as they can understanding visual ones.

Another experiment in this series investigated the relationship between preschool children's performance on A–R tasks and Level 2 perspective-taking (PT) tasks. In this study, five Level 2 Color PT tasks (e.g., an object of one color was placed on the experimenter's side of a blue filter presented perpendicularly to the table and parallel to the child, and the child was asked how the object looked to the experimenter and to him/herself), and five Shape PT tasks (e.g., a straight object was held behind cylindrical bottle of fluid and child asked how the object looked to experimenter and child) were administered to the subjects, along with five Color A-R and Shape A–R tasks. The results indicated that the ability to make appearance–reality distinctions was highly correlated with Level 2 perceptual perspective-taking abilities. Thus, concluded Flavell et al. (1986), these skills appear to share common knowledge and skill components. Flavell et al. (1986) believed that this common set of skills and knowledge may mediate successful performance on A–R and Level 2 perspective-taking tasks, and cause the tasks to be mastered at about the same developmental stage.

Gopnik and Astington (1988) investigated young children's (3, 4, and 5 year olds) ability to understand representational change (i.e., that although objects remain the same, one's representations about the objects can change), and its relation to the understanding of false-belief and appearance–reality distinction. In their first experiment, Gopnik and Astington (1988) presented each subject with the Smarties task and the Rock task described earlier, with the object shown in its deceptive form first. After the object was returned to its deceptive form (i.e., the Smarties box with the pencil inside, closed, and sitting on the table, and the "rock" placed on the table across from the subject), the children were asked a representational change question (i.e., "When you first saw the box, before we opened it, what did you think was inside it? Did you think there were pencils inside it or did you think there were Smarties inside it?"), a false-belief question (i.e., "If X sees the box all closed up like this, what will [s]he think is inside it? Will [s]he think there are pencils inside it or will [s]he think there are smarties inside it?"), and a pair of appearance–reality questions (i.e., "Does it look like this box has pencils in it or does it look like it has smarties in it?, and, What's really inside this box? Are there really pencils inside it, or are there really smarties inside it?"). On the representational change task, Gopnik and Astington found a significant age effect, with performance increasing across the three age groups ($F(2,15) = 5.38$, $p < .05$). Specifically, less than one-half of the three year old group answered the representational change question correctly, while more than two-thirds of the five year old group answered the question correctly. Difficulty on the representational change question was also related to difficulty in understanding false-belief and appearance–reality distinction. Significant correlations were found between performance on the change question and the appearance–reality questions ($r = .46$, $p < .01$), and between performance on the false-belief and appearance–reality questions ($r = .49$, $p < .01$).

In order to control for a number of perceived methodological flaws present in the first experiment, Gopnik and Astington (1988) conducted a second experiment in which the form of the change question was simplified (i.e., "When you first saw the [Smarties] box, all closed up like this, what did you think was inside it?"). Additional sets of materials were used as well, to see if children would show the same difficulties across materials, or if some materials would be more difficult than others.

The results indicated that the particular syntactic form of the questions had no effect, but there was a significant age effect and task material effect ($F(2,30) = 6.39, p < .01$ and $F(4,120) = 5.85, p < .001$ respectively). Like Flavell et al. (1983), and Flavell et al. (1986), Gopnik and Astington (1988) found that realism errors predominated on the object Identity tasks. Although the children's performance on the rock and Smarties task was comparable to that of Experiment One, a book task (in which the child had a restricted view of "bunny ears", but when the page was turned it became apparent that the "ears" were petals of a flower) and a cat task (in which a green cat was covered by a pink transparency which made it look black) appeared to be easier. The Doll task, however, (in which two rag dolls that could be put together and covered with one dress making it look like one doll) appeared to be more difficult than the Rock and Smarties tasks. Thus, Gopnik and Astington's (1988) research provided additional support for the developmental perspective of representational change, false-belief and appearance–reality distinction.

All of the A–R studies discussed above are particularly valuable in that they have taken into account the possibility that the syntax and semantics of the questions asked affected how the children responded. Accordingly, the investigators adjusted questioning techniques on subsequent experiments, included orientation, pretraining, and feedback tasks and counterbalanced questions to control for this possibility. Despite these procedures, the results were consistent. Most three year olds had little or no conception of the A–R distinction, and this ability improved with age. Four and five year olds had a fairly good understanding of the concept, while six and seven year olds had developed the skill. In all of the studies, a consistent pattern of errors was found in the children's responses. A–R tasks involving object properties evoked phenomenism errors, while A–R tasks that involved object identities elicited realism errors.

Instruction of Autistic Children

Over the years, many educational techniques have been tried with autistic children to remedy different deficits. Specific techniques differ depending on the individual child's needs and deficits as outlined at the beginning of this chapter, and depending on theoretical orientation (i.e., psychogenic versus biological etiologies of autism).

Contributors to the autism literature currently suggest that behavioral methods of teaching are the most effective with this population (Baron, Groden, & Cautela, 1988;

Lovaas, 1977, 1981). The incorporation of such techniques as positive reinforcement for successive approximations, prompting hierarchies (i.e., verbal, partial or full prompts), and shaping and fading procedures into educational programs for autistic children have resulted in language acquisition (Lovaas, 1977), functional and academic skills, and more appropriate social behavior (Schopler, Reichler, & Lansing, 1980).

Most researchers agree that a high degree of structure is an essential aspect of the learning environment for autistic children (Bartak & Pickering, 1976; Gallagher & Wiegerink, 1976; Baron et al., 1988; Lovaas, 1977, 1981; Schopler et al., 1980; Wing, 1980). Schopler et al. (1980) noted space, time and duration of teaching sessions as important aspects to be considered in designing educational programs for autistic children.

Space

Many autistic children appear to be overwhelmed when confronted with multiple stimuli (Hermelin & O'Conner, 1970). Therefore, the room in which teaching is conducted should be quiet, and free from both visual and auditory distractions. It is also best if teaching can be conducted in the same room to make it easier for those children who have difficulty adjusting to new surroundings. If the children have less to adapt to, then instructional benefits can be maximized.

Time

Time of teaching is also important to consider when teaching autistic children. Some children find it easier to adapt to situations that have become predictable and part of a routine. Therefore, instruction may be more effective if a particular lesson consistently takes place at a certain time of day.

Duration

Autistic children vary widely in terms of their attention span, but in general, short, frequent instructional sessions tend to be more effective than long, less frequent ones. Schopler et al. (1980) noted that each teaching session should contain a number of activities, and that the length and pace of each activity may vary from day to day depending on the children's ability to attend, and their speed of working on a given day.

All of these considerations led the current investigator to select Direct Instruction (DI), a behaviorally based approach, as the method of instruction in the current study. Direct Instruction is a highly structured teaching approach that has been found to be beneficial in teaching a variety of children with special needs (Englert, 1984; Gerston, 1985), including those with autism (O'Neill & Dunlap, 1989).

Direct Instruction

The Direct Instructional method views the analysis of cognitive learning as a combination of three separate analyses. The first analysis is that of behavior, and attempts to explain how the environment affects learning. This analysis provides the basis for determining the method of instruction, apart from the actual material being taught. That is, the analysis of how the learner reacts to certain reinforcement, chaining, and shaping procedures will determine the most efficient method for using those procedures in presenting instructions, questions and examples (Carnine & Becker, 1982; Moore, 1986). This form of analysis is called the Response Locus Analysis (Moore, 1986).

The second analysis is that of knowledge systems, in which the three knowledge types (i.e., facts, correlation and cause/effect) are categorized to determine the most effective method of presenting each category to the learner. The final analysis is that of communication, which attempts to determine the most effective way of presenting the knowledge categories (Moore, 1986). The analysis of knowledge systems and communication make up what is known as the Stimulus–Locus analysis. These analyses emphasize the antecedent events of a learning situation, along with the selection of content for instruction, and the most effective method of presentation of the content (Moore, 1986).

Becker and Carnine (1980) identified six shifts that characterize the instructional sequence of DI. The first of these is the shift from overtized to covertized problem-solving strategies, in which the teacher begins instruction by making each step of the problem-solving process explicit, prompting the learner to follow the same steps as the teacher in solving the problem. As the strategy becomes covertized, the teacher no longer requires a response at every step of the problem-solving process. Eventually, only the answer may be overt.

The second shift is from simplified contexts to complex contexts. The saliency of the relevant stimulus features is maximized when a new discrimination is introduced, but as these discriminations are mastered, more complex contexts are introduced in which irrelevant detail is introduced into the stimulus.

The fading of prompts is the third shift. When instruction begins, the teacher may include various prompts that focus the student's attention on relevant stimulus features. As the student continues to respond correctly, these prompts are gradually faded. For example, in teaching a child how to write cursively the lowercase letter "a", the teacher may emphasize the fact, through pointing and/or verbalization, that the letter must be completely closed in order to be an "a" (and not a "c"). As the child becomes more adept at closing the letter, the prompts may be reduced, and then eliminated altogether.

The fourth shift is from massed practice on new skills to distributed practice. That is, when a skill is being introduced to a child, many trials and examples are given. As the child gains mastery of the skill, the skill is then interspersed with other skills and is presented more periodically. The early massed practice is designed to assist in mastery learning, while distributed practice assists in retention.

The transition from immediate to delayed feedback constitutes the fifth shift, while the final shift is from the teacher as a source of information to the learner as a source of information. As the learner becomes more capable, and his/her repertoire of skills increases, the teacher becomes more of a guide on how to use information, than a provider of new information. The student thus learns to apply previously learned skills in new problem-solving situations.

Concept Generalization

Because the learning of any concept is limited by the number of examples that can be feasibly presented during instructional sessions, the generalization of learned concepts is an important consideration. The study of generalization "focuses on the analysis of stimulus qualities, the structural basis for generalisation [sic], that lies in the potential teaching examples" (Carnine & Becker,1982). Carnine and Becker (1982) have specified a number of structural conditions that the set of teaching examples must meet if generalization is to occur.

> 1. The communication must present a set of examples that are the same with respect to one and only one distinguishing quality (the quality that is to serve as the basis for generalisation). (This may also be a set of qualities.)
> 2. The communication must provide two signals, one for every example that possesses the quality that is to be generalised, the second to signal every example that does not have this quality. (This is a conditioned variation of differential reinforcement.)
> 3. The communication must demonstrate a range of variation for the positive examples (to induce a rule that is appropriate for classifying new examples on the basis of sameness). (The basis for interpolation.)
> 4. The communication must show the limits of permissible variation by presenting negative examples. (The basis for extrapolation.)
> 5. The communication must provide a test of generalisation that involves new examples that fall within the range of variation demonstrated earlier. (p. 251)

Stokes and Baer (1977) stated that if generalization is a desired outcome of instruction (as it usually is), then it must be actively programmed for in the instructional process. Direct

Instruction acknowledges this need, and builds generalization processes into each teaching sequence.

Direct Instruction and Autistic Children

Research has shown that a number of DI features have particular relevance in the teaching of autistic children. For example, Direct Instruction involves bringing responses under stimulus control, the use of varied tasks within a teaching session, rather than repetition of a single task, and rapid pacing of task presentation. These same techniques have also been found to be effective with autistic children (Dunlap, 1984; Dunlap & Koegel, 1980; Koegel, Dunlap, & Dyer, 1980; Rincover, 1978; Rincover & Koegel, 1975).

For example, Rincover and Koegel (1975) found that low-functioning autistic children who learned to respond to a request in a treatment setting, failed to transfer learning to a novel setting. It was determined that the children were responding to incidental stimuli in the training environment rather than the stimulus of relevance. When the irrelevant stimuli were introduced to the novel setting, however, the children's responding increased substantially.

Direct Instruction methods concentrate on ensuring that the child is responding to the relevant stimulus. The stimulus-locus analysis "involves the logic of *ruling out all the possibilities but the one* to be conveyed through a teaching communication" (Engelmann & Carnine, 1982, p. 8).

The use of varied tasks has also been shown to be effective in teaching autistic children. In an experiment investigating the relative effectiveness of constant versus varied task conditions, Dunlap and Koegel (1980) found that when learning tasks were varied, autistic children produced substantial gains in correct responding and maintained these gains regardless of the number of trials or sessions that were conducted. The subjects also scored higher on dimensions of enthusiasm, interest, happiness, and general behavior, during the varied task condition, than during the constant task condition.

In a similar experiment, in which the effectiveness of a constant task condition was compared with varied-acquisition-task and varied-with-maintenance-task conditions in learning 10 tasks, Dunlap (1984) also found that significantly more efficient learning took place under the varied-maintenance task condition, with no consistent differences being noted between the other two conditions. Like the Dunlap and Koegel (1980) experiment, ratings by observers indicated that the children showed the most positive affect in the varied-maintenance task condition.

The varied-maintenance task condition's correlate in Direct Instruction is the shift from massed to distributed practice. This shift maximizes both interest, by reducing frustration generated by constantly encountering new and difficult material, and retention, through periodic repetition.

Research has shown that teachers who maintain a rapid pace of presentation when using Direct Instruction techniques, elicit significantly more correct responses from children than do teachers whose pace of presentation is slower (Englert, 1984). Similarly, instructional studies with autistic children (e.g., Koegel, Dunlap, & Dyer, 1980) demonstrated that inter-trial intervals of short duration (maximum range of 1 to 4 seconds) resulted in higher levels of correct responding than did long (maximum range of 4 to 26 seconds) inter-trial intervals.

Reinforcement, whether intrinsic or extrinsic, plays an important role in children's learning, in general, and autistic children's learning, in particular. When children are unable to complete tasks correctly, their motivation to respond to instructional tasks at all, is greatly diminished (Dunlap, Koegel, & Egel, 1979; Koegel & Egel, 1979). Because of their severe learning handicaps, autistic children may experience lower levels of reinforcement, consequently lowering their attempts to respond to the presented tasks (Rincover, Cook, Peoples, & Packard, 1979). Rincover, Newsom, Lovaas, and Koegel (1977), and Rincover et al. (1979) concluded that educational procedures for autistic children must maximize their opportunities for correct responding and receipt of reinforcers.

Direct Instruction provides a virtually errorless learning method whereby answering a question incorrectly is extremely difficult, since the teacher initially orients the student to the task, models the correct answer and subsequently leads the student to the answer, leaving very few options as to what the correct answer is. In most Direct Instruction teaching sequences, a range of both positive and negative examples of the concept being taught is presented to the learner. The positive examples serve to reveal the defining features that are the same in each concept example. The negative examples, on the other hand, serve to indicate the borderline of the concept. The negative examples specify the qualities of the concept that are *not* the same (Kameenui & Simmons, 1990). For example, in teaching the concept "farther apart", the teacher tells the student what will be demonstrated, and then asks the student to verbalize what will be demonstrated. The teacher then models three positive examples of "farther apart," verbalizing the concept each time the teacher's hands are moved farther apart (e.g., "My hands are farther apart."). Subsequent to the positive examples, the teacher then demonstrates negative examples in which the teacher's hands are not moved farther apart (i.e., the hands are moved closer together), and verbalizes the example (e.g., "My hands are not farther apart."). Finally, the student is tested on the sequence and the student is required to respond to the teacher's presented positive and negative examples of farther apart. When an error occurs, it is corrected immediately, and an example is repeated. If an error persists, the entire sequence is re-taught within the same teaching session (Kameenui & Simmons, 1990).

Because the concept to be taught in the current study does not fit into a specific Direct Instruction knowledge category, with an established method of presentation, the described DI procedures have been adapted to suit the needs of the proposed study. Thus, the essential features of DI, including task analysis, the establishment of stimulus control, and rapid pacing have been retained, but the actual presentation sequence has been adjusted. The presentation method that will be used in the study is explained fully in the next chapter.

To summarize, the effectiveness of many features of Direct Instruction, including the establishment of stimulus control, rapid pacing, the use of massed and distributed practice, and prompts, has been demonstrated in research with both autistic and nonautistic children. Because of its effectiveness with autistic children, and the relative ease of quantifying performance for research purposes when using Direct Instruction, this approach was selected for teaching the appearance–reality distinction.

Summary and Limitations of Current Research

In summary, investigators studying autistic children have repeatedly demonstrated that these children experience a specific developmental delay in their acquisition of a theory of mind. In nonhandicapped children, the precursors to theory of mind emerge in the second year of life in the form of pretend, or symbolic play, perceptual perspective taking and joint-attentional gestures. While autistic children show no impairments in perceptual perspective taking tasks, they do display significant impairments when it comes to symbolic play, conceptual perspective taking tasks, and understanding some causes of emotions. Most autistic children with approximate CAs of 13.0 are unable to attribute first-order beliefs or correctly make appearance–reality distinctions, abilities that are fairly well developed at about age four in normal children (Leslie, 1987, 1988). Autistic children who can attribute first-order beliefs, are unable to correctly attribute second-order beliefs, an ability evident in normal children at six to seven years of age (Perner & Wimmer, 1985).

The experiments including autistic children that were described earlier lend strong support to the theory that autistic children manifest significant deficits in their acquisition of a theory of mind. The studies are particularly significant given the researchers' careful methods of subject selection: the use of normal and mentally retarded control groups, IQ matching, and the use of fairly high functioning autistic children, all of which assisted investigators in concluding that the observed deficits are attributable specifically to autism. This stringent inclusion criterion is considered essential in autism research, but is often ignored (Kistner & Robbins, 1986). However, most of the published experiments conducted with autistic children on this topic have been conducted by the same researchers, with children from the same schools. Replication of these results by other researchers and with other autistic children is needed to enhance the robustness of the theory.

Researchers conducting theory of mind studies with normal children have examined the possible role of the linguistic and syntactical form of the questions (Gopnik & Astington, 1988), the influence of specific task materials and requirements (Flavell et al., 1986; Gopnik & Astington, 1988), and the influence of stimulus salience (Flavell et al., 1983). These researchers have also investigated the influence of young children's failure to retain essential facts, failure to understand the normal expectations which give rise to false belief, and pragmatic misinterpretation of the test questions (Perner et al., 1987). It may be wrongly assumed that autistic children's performance would be similar to that of young normal children who fail to employ a theory of mind, or who employ a lower level of theory of mind. It is possible that high functioning autistic children (i.e., those who are able to correctly answer first-order false-belief attribution questions) may exhibit situation-specific abilities. That is, some autistic children may be able to perform as accurately as normal children in some aspects of theory of mind, and unable to understand other aspects.

A particular weakness that has been noted in most of the published appearance–reality distinction studies (e.g., Baron-Cohen, 1989a; Flavell et al., 1983; Flavell et al., 1986; Gopnik & Astington, 1988) is the wording of the appearance–reality questions. In each case, the subject has been asked, "What does (the object) *look* like?", and "What is it *really*?". Both of these questions imply a previously acquired understanding of the lexical terms "looks like" and "really." The researchers have acknowledged this difficulty, but do not seem to have dealt with the problem adequately.

For example, Baron-Cohen (1989a) familiarized subjects with the terms "looks like" and "really" by closing the curtains, turning off the light in the room, and telling the child, "Look. We can close the curtains and turn off the light. Now the room is dark and it looks like it's the night. But really it's day time." The child was then asked, "Does it look like day or night?", and, "What is it really?" Corrective feedback was given to those subjects requiring it, but the question remains whether or not the concept was truly understood. Even if the subject was able to answer the Appearance question correctly (a 50 percent chance of being right), the subject may have automatically chosen the second option in answering the second question, simply because a different answer is implied by the juxtaposition of the questions.

In Baron-Cohen's (1989a) appearance–reality study, the subjects were given a Memory pretest prior to administration of the appearance–reality tasks, to ensure that the subjects knew that the colored filter did not permanently alter the real color of the object, and that the subjects remembered the original color of the object. For this pretest a white sheet of the paper's real color was altered by placing an orange filter on top of it. The subjects were then asked, "When I take this filter away, will the paper look white or orange?" Baron-Cohen (1989a) and others (Flavell et al., 1983; Flavell et al., 1986) noted that passing such pretests is a prerequisite for passing appearance–reality tasks that involve transformations. However, it appears to this author that the pretest *was* the appearance–reality distinction. By asking the subjects what color the paper will be when the filter is taken away, the experimenter is implicitly asking the subjects what the object's true color is, perhaps in simpler terms than the traditional A–R question format. All of Baron-Cohen's autistic subjects were able to pass this pretest, perhaps suggesting that autistic children are better able to understand the requirements of the A–R task when the task is presented in the manner of Baron-Cohen's (1989a) pretest.

The procedures used in studies by Flavell et al. (1983), and Flavell et al. (1986) may have also posed linguistic problems. For example, in one study (Experiment 1, Flavell et al., 1986), the Appearance question was, "When you look at the seal with your eyes right now, does it look pink or does it look green?", and the Reality question was, "What color is the seal really and truly? Is it really and truly pink or really and truly green?" Such terminology ("really and truly") does not ensure that the subject understands the concept any better than understanding the term "really" in isolation. Similarly, emphasizing "really" by asking the subject, "What is this *really, really*? Is it *really, really* a (rock) or *really, really* (a piece of sponge)?", as did Flavell et al., (1983), also does not ensure that the subject understands the meaning of "really." Thus, the adequacy of the procedures of published studies in demonstrating the concepts of "looks like" and "really" is questionable.

Flavell, Flavell, and Green (1987), however, were able to circumvent the lexical problem by having the subjects complete the A–R task nonverbally. In Flavell et al.'s (1987) study, the subject observed the experimenter place a white card under a blue filter so that the card looked blue. With the card still under the filter, the experimenter removed a pre-cut piece from the card, put the piece into a closed hand, withdrew the hand from the filter, and placed that piece and a blue piece of paper of the same dimensions on the table in front of the subject. The child was asked, "Which is the piece I just took out of the card?" This procedure is capable of eliciting the appearance–reality distinction, but does not require an understanding of the terms "looks like" or "really". Flavell et al., (1987) found that three year old subjects were incapable of making the appearance–reality distinction when the problem was presented in this format as well, thus verifying the subjects' inability to understand the appearance–reality concept, rather than the lexical terms of "looks like", and "really." Hence, the linguistic form of the questions may not play a role in children's comprehension of the task. However, its possible contribution to confusion must be considered.

Methodological weaknesses in those studies that attempted to teach the appearance–reality distinction (i.e., Flavell et al., 1983; Flavell et al., 1986) may also account for these researchers' lack of success. A number of reasons may account for the inadequacy of attempts at teaching the A-R distinction. Firstly, the training the children underwent in the various experiments may have been of too short duration for the acquisition of this type of conceptual understanding. Training on the tasks never exceeded five to seven minutes. Although Braine and Shanks (1965) were able to elicit correct answers from subjects with their training procedure, the subjects who were successful in their study were older than those in the Flavell et al. (1983) and Flavell et al. (1986) studies, thus making comparison between the studies difficult.

Secondly, it is possible that the training procedures involved conceptual leaps which the children were unable to follow. For example, in both the Flavell et al. (1983) and the Flavell et al. (1986) studies, the objects used in training appeared *completely* as either one color (or size) or the other (e.g., the white lamb puppet was either white [real color] or red [apparent color when behind filter]; the piece of paper was either little [seen without using magnifying glass] or big [seen with magnifying glass]). At no time in the training procedures were gradations in the objects' color or size used. That is, the child did not see the white puppet slowly become red as it was put behind the filter, or the piece of paper slowly become big as a magnifying glass was placed over it. The object was either behind the filter/magnifying glass, or it was not. Therefore, to the child's eyes, (and perhaps mind), the object was either red or white, big or small, and nothing in between. Nor was the child shown how, at various points of the change process, the true color/size of the puppet/paper remained unchanged. Given young children's tendency to constantly report perceptual appearance in object property tasks (Braine & Shanks, 1965; Flavell, 1988), it may be that young children are unable to visualize the intermediate steps of the change process. The object "magically" becomes a different color/size, in the case of object properties, and a different object completely in the case of object identity.

A third reason why training in these studies may have been unsuccessful, is that the directions might have been too complex for three year old children to understand. Flavell et al.'s (1983) feedback and pretraining procedures consisted of complex sentence structures, and although the aim of the pretraining was to demonstrate the meaning of the words "looks like" and "really", the words were emphasized and repeated, rather than explained. Thus, the meaning of "looks like" and "really" may have remained ambiguous to the children. In terms of instructional implications, then, it is possible that children who undergo A–R instruction that is of longer duration, is presented in such a manner as to minimize the number of cognitive leaps that may be necessary, and in which the meaning of the critical words is more clearly explained (or demonstrated), may be more successful in learning the distinction.

In addition, instructional techniques have only been attempted with nonhandicapped children. It remains unknown if autistic children would benefit from instruction. The current study of appearance–reality distinction in autistic children will further examine these children's ability to understand the appearance–reality distinction. It will attempt to teach the distinction to children unable to pass the type of appearance–reality tasks described in the studies discussed above (i.e., Color and Size A–R tasks). The appearance–reality dimension of theory of mind was selected specifically because only one published experiment to date has dealt with this aspect at all, with respect to autistic children.

The focus on instruction of the appearance–reality distinction brings up the issue of whether the forced acquisition of this ability is even desirable in normally developing children. From a research perspective, it is of interest to determine if young normal children's theory of mind abilities represent true developmental distinctions, or are of artificial artifact. From a practical perspective, it can be argued that normally developing children will acquire these abilities in due time without intervention. However, for autistic children experiencing significant delays in the acquisition of theory of mind, and theory of mind's postulated connection to autistic children's social and communication deficits, knowing whether instruction of various theory of mind abilities can be successful is of great importance.

CHAPTER THREE

METHODOLOGY

Inclusion Criteria

A number of criteria were established to determine children's eligibility for the study. First, potential subjects had to pass a memory test to ensure that they were able to remember an object's original appearance for a period corresponding to the duration of the experimental task. Secondly, to be included in the study, the subjects were required to fail three appearance-reality distinction tasks (Color, Size, and Object Identity). This criterion was established to ensure that the selected children did not already possess the appearance-reality distinction concept. These inclusion tasks are described below. The Peabody Picture Vocabulary Test-Revised (PPVT–R, Dunn & Dunn, 1981) and the Leiter International Performance Scale (LIPS, Leiter, 1979) were also administered to obtain respectively, a receptive vocabulary age-equivalent score and a nonverbal mental age score. These tests are also described below.

In addition to these standardized tests, a language sample was obtained from each subject, with the aid of a speech pathologist, to determine each subject's mean length of utterance (MLU). Because of the inherent linguistic difficulties of appearance-reality distinction tasks, as mentioned in the previous chapter, the investigator felt that the MLU information was essential for a comprehensive evaluation and interpretation of this study's results. This information will also assist the reader in interpreting the results of this study within the wider context of expressive and receptive language development, a context that has been largely ignored in the theory of mind research conducted with autistic children.

The procedures presented in this chapter explain the *initial* standardized method used by the investigator. However, minor individual variations were made in the procedure as the investigator discovered particular patterns of responding among the subjects. In addition, if specific subjects did something novel with the objects used in the study, the investigator followed the sujects' lead and questioned them accordingly.

Rather than detracting from the methodological rigor of the study, the investigator believed that valuable additional information was gained by this individualization. Any individual alterations that were made in the presentation format are discussed in the next chapter as each subject's results are discussed.

Subjects

Ten autistic children were initially screened for participation in the study. However, five of these children were excluded because they did not meet the inclusion criteria. Specifically, two of the excluded students passed the initial theory of mind appearance–

reality tasks, thus demonstrating that they already possessed the abilities targeted for instruction. The other children excluded from the study did not possess the attention or language skills necessary for participation.

The final subject sample included one girl and four boys, all attending public school in Edmonton, Alberta. Three of the children selected had been assessed at a local hospital and were diagnosed as autistic, while the remaining two children had been assessed by special education consultants of a local school board. All subjects met the school board's criteria for classification as autistic. The chronological age of the children ranged from 9.4 to 12.1, while Receptive Vocabulary Age Equivalent Scores as measured by the PPVT–R (Dunn & Dunn, 1981), ranged from 3.5 to 4.7, and nonverbal Mental Age scores as measured by the LIPS (Leiter, 1979) ranged from 4.5 to 6.10. Mean length of utterances (MLU) of the subjects ranged from 1.9 to 4.37. Table 3.1 presents background data on each subject.

Table 3.1

Subject Background Data

Subject	Chronological Age	Receptive Vocabulary Age Equivalent	Mean Length of Utterance	Nonverbal Mental Age	School Placement
Nicky	9.4	4.5	4.17	5.3	-primarily in segregated class for autistic students but integrated in grade 1 class for a few subjects
David	11.7	4.7	4.37	6.10	-completely integrated with aide assistance in grade 3/4 class
Brian	9.8	5.6	1.89	4.5	-completely integrated with aide assistance in grade 4 class
Bruce	10.9	3.8	2.5	5.1	-integrated with aide assistance about 50% of the time in grade 3/4 class with remaining time spent in segregated class for autistic students
Steve	12.1	3.5	2.6	5.4	-in segregated class for autistic students

Although it was originally intended that the subjects should have receptive vocabulary age equivalents and nonverbal mental ages of at least 7.0, time constraints and subject availability made these criteria impossible to achieve. However, all but two of the subjects achieved receptive vocabulary age equivalents above 4.0, while all subjects scored at or above the 4.6 level on the test of nonverbal mental age. Thus, most of the subjects were at the mental level at which non-autistic children are able to correctly complete the appearance–reality tasks. Therefore, the *possibility* of success in learning the appearance-reality distinction seemed feasible (Baron-Cohen, 1989b; Leslie, 1987, 1988). Interpretations of the results of the following research, however, must be made cautiously, given the subjects' limited verbal capabilities.

Setting

The administration of the subject selection tasks, the collection of baseline data and the training procedure, took place in a conference room within each child's school, a room in which privacy was assured and which, for the most part, was available for the duration of the project. Each room was free from auditory and visual distractions. The tables and chairs used during the study were appropriate to the respective sizes of the participating children. Hence, for the smaller children, a small table and chairs were used, while for the taller children, a standard sized table was used. Throughout the study, only the materials that were being used at any given time were available to the children. Materials not being used remained out of the children's sight, placed either on the floor by the investigator's feet, or in a box on the floor. All children were tested and taught on an individual basis.

Inclusion Materials and Procedure

Memory Test

For the memory test, three objects, independent of the study (a red Lego block; a green, plastic, two dimensional elephant, approximately 7.5 cm x 6 cm; and a three dimensional blue dog made of rubber, approximately 5 cm x 2 cm x 3 cm) were used. The first object was placed on the table in front of the child and the experimenter, while pointing to the object, asked the child, "What color is this?" After the child correctly identified the color of the object, the object was hidden from view by placing a cardboard screen between the object and the subject. The experimenter then asked the subject the color of the object. The task was repeated for each object in a random order.

The use of the cardboard screen proved difficult for Nicky and Brian, since, despite verbal prompting, they repeatedly reported the color of the screen, rather than the color of the object behind the screen. Thus, the presentation format was modified for these subjects, and each object was hidden in the experimenter's hands rather than behind the screen. All subjects were then able to correctly identify each object's color. To be included in the study, a child needed to correctly identify the colors of each object when the object concerned was not in view on each of two trials. An example of the scoring form for the memory test is presented in Appendix B.

Appearance-Reality Inclusion Tasks

The Color and Size appearance–reality tasks used in the subject selection procedure corresponded to the type of A–R tasks that were taught during intervention, and constituted the pre- and post-test measures. That is, similar, but not identical, materials and task requirements characterized both the inclusion tasks and the instructional tasks. In addition, an Object Identity task was administered during pre- and post-testing, and once during instruction, at the point where each subject reached criterion on either the Size or Color tasks (whichever reached criterion first). However, the object identity aspect of the appearance–reality distinction was not taught directly. Since the materials used during pre– and post–testing were not identical to those used during instruction, the obtained pre– and post–test scores also served as a measure of generalization of the appearance–reality distinction.

The order of administration of the tasks, as well as the order of the Appearance/Reality questions, were randomized across subjects to control for any potential order effects. A description of the procedures that were followed during appearance–reality pre- and post-testing is presented below. All of the appearance–reality pre/post test procedures and the Object Identity tasks were adapted from Baron-Cohen (1989a), Flavell et al. (1983), and Flavell et al. (1986). The post-test for a given type of task (i.e., Color or Size) was administered on the last day of instruction of that task type, or as close to the last instructional session as possible.

With the exception of the Koala task, all Color and Size tasks were presented using a bookstand as a presentation surface during all testing and instruction. As well, a piece of black construction paper was used as an easel for all color tasks except the Koala task. The purpose of this was to minimize the color distortion of the area surrounding the object.

Color

Materials for the Color A-R pre- and post-test task included a pair of plastic yellow scissors (approximately 12.5 cm x 6 cm) and a red transparency (filter). When the filter was placed over the scissors, the scissors appeared orange.

The Color A-R pretest proceeded as follows. The experimenter placed the pair of yellow scissors on the easel and asked the subject, "What color are these?" This verified the child's ability to identify colors. The red transparency was then placed over the scissors, and the child was asked the A-R pair of questions: "What color do they *look*?" (Appearance question), and "What color are they *really*?" (Reality question). Despite the methodological problem of using this questioning format, as noted in the previous chapter, it was retained during this study since this has been the most common format used in the published studies.

Size

The Size A-R inclusion task used two blue buttons of different sizes, one approximately 2.2 cm in diameter, and the other approximately 1.7 cm in diameter, affixed side-by-side on a piece of yellow poster board and placed on the bookstand. The pretest began with the experimenter asking a child to identify which object (button) was bigger, thus determining the subject's ability to discriminate size. In the description that follows, the reader is directed to Figure 3.1. A sheet of plexiglass (A), on which a small magnifying glass (B) and an equal sized piece of clear plexiglass (C) were mounted, was positioned in front of the bookstand in such a manner that the smaller button (D) was magnified and appeared larger than the objectively larger button (E).

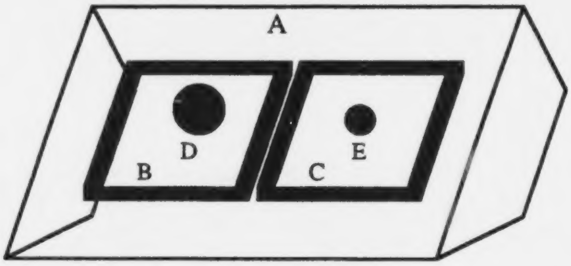


Figure 3.1. Size task apparatus.

The experimenter then asked the Appearance question, "Which one looks bigger?" and finally the Reality question, "Which one is *really* bigger?" The Appearance and Reality questions were sometimes phrased as, "Point to the one that *looks* bigger", and "Point to the one that's *really* bigger." In the course of the study the experimenter found that Brian appeared not to know what response was required of him when he was asked the Appearance and Reality questions in the former format. Questions asked in this format resulted in echolalic responses. However, a pointing response was established when the latter question format was employed and echolalic responses were eliminated.

To maintain consistency of procedure across the subjects, the latter question format then was introduced to the other subjects as well. This difference in question format did not appear to alter the kind of responses the subjects made. That is, regardless of the question format, the subjects responses remained consistent.

Object Identity

The Object Identity pre-test task used a realistic looking "orange" made of plastic. The task was introduced to the subject as follows: "Now I'm going to show you something and ask you some questions about it. Are you ready?" The experimenter then removed a cardboard screen to reveal an orange placed on a cloth placemat. The screen prevented the possibility of the child discovering the object's true nature by observing its movement, while the mat prevented the orange from rolling around and making a sound uncharacteristic of an orange. The subject was asked, "What is this?" The response, "orange" confirmed the object's realistic appearance. The subject was then allowed to handle the object to discover its plastic ("fake") quality. The subject was then asked a probe question, "What is it made of?", to ensure that the child realized that the object was not real. This question proved to be very difficult for all of the subjects during either the pre-or post-testing, or both. For example, one subject (Bruce) replied, "lemonade," while Steve and Brian did not respond to the question at all. If a child was unable to correctly answer this question, the experimenter then asked, "Is this a real orange or a pretend orange?" Nicky, Brian, and Steve were then able to correctly identify it as a pretend orange. Upon review of the subjects' responses the investigator realized that subjects may have been responding to the last word (pretend) rather than the objects' properties. Hence, during post-testing and administration of the Rock task the subjects were asked if the object was pretend or real. Bruce, however, apparently still did not understand the question. The experimenter asked him, "Can you peel *this* orange? Can you eat *this* orange?" Even with this amount of direction, Bruce still did not appear to understand. The Object Identity testing was discontinued with Bruce. For the remaining subjects, the orange was returned to its deceptive appearance (i.e., on the table across from the subject), and the subject was asked the Appearance question ("When you first saw this, over here like this [experimenter pointed to the orange], what did you think it was?"), and finally the Reality question ("What is it *really*?" [experimenter again points to the orange]). Again, none of the subjects appeared to understand these questions. Because of the difficulty encountered in using the plastic orange, the same task was attempted with imitation carnation flowers, and a plastic egg that looked like a broken raw egg sitting on a plate. The same difficulties were encountered with these objects as with the "orange." Despite the subjects' difficulty with these tasks, the second Object Identity task (using a "rock" made of foam) was administered after criterion was reached on the Size or Color task regardless of the subjects' performance on the Object Identity pretest. The purpose of the administration of the Rock task was to determine the effect, if any, on the understanding of an Object Identity task, after understanding on a similar task (i.e., Color or Size task) had been achieved.

Peabody Picture Vocabulary Test–Revised

Boucher (1989) has criticized the use of one word vocabulary tests in theory of mind studies, because they do not assess the understanding of complex grammatical questions required for success on the tasks. Boucher (1989) observed that some autistic children develop large one word vocabularies, but that their grammatical competence may develop at a slower pace. Despite this valid criticism, the PPVT–R (Form M) was chosen for this study, because of its similarity to the British Picture Vocabulary Scale (used in the British theory of mind studies), and the necessity of comparing this study to the British studies since no comparable Canadian studies have been conducted.

The PPVT-R (Forms L and M) is a verbally presented, multiple-choice test designed to measure the receptive vocabulary ability of children and adults. The subject is showed pages, each with four pictures, and asked to point to the picture that best defines a particular word the examiner says. Testing is continued until the subject gets six errors within eight items.

Split-half reliability coefficients on the PPVT-R for ages 2.6 through 18.0 years range from .61 at the 7.0 year level to .86 at the 12 year level (median r = .81). Stability of the PPVT (the unrevised edition) was found to be particularly high for mentally handicapped children (Sattler, 1982) but similar studies have not yet been conducted with the PPVT–R.

Information on the validity of the PPVT–R is not included in the test manual. However, concurrent validity studies between the PPVT and intelligence tests have yielded median validity coefficients in the .60s with a range from the .20s to the .90s (Sattler, 1982).

Although researchers using the British Picture Vocabulary Scale (a British standardization of the PPVT–R) have considered the test a measure of verbal mental age in their studies, a claim that is unjustified given the limited nature of the test, it was considered a measure of receptive vocabulary in the present study.

For the current study, the PPVT–R was administered and scored according to the manual. A receptive vocabulary age-equivalent was obtained.

Leiter International Performance Scale

The Leiter International Performance Scale (Leiter, 1979) is a test of nonverbal intelligence that has been widely used with various handicapped populations. This test has also been used extensively in theory of mind studies with autistic children to yield a nonverbal mental age (Baron-Cohen, 1989d; Baron-Cohen et al., 1985, 1986).

The Leiter International Performance Scale (LIPS) is a 54 item nonverbal test of intelligence considered appropriate for testing children from two to 18 years of age having hearing impairments, expressive and/or receptive language disabilities, mental retardation, cerebral palsy, or who are non-English speaking. For each item there is a paper strip and a set of blocks. The subject is required to place the blocks into the appropriate stalls on a wooden response frame. The tasks range from perceptual items involving the matching of colors and shapes, block design and visual closure, to highly abstract conceptual items involving deduction of categories and underlying relationships, numerical processes, and visual-spatial relationships. Given the range of tasks included in the test, the LIPS is considered to tap a wide range of intellectual abilities (Matey, 1985). Items are scored objectively as either pass or fail, and months of credit for each item passed are summed to derive a mental age score, and an intelligence quotient.

Many studies on the reliability and validity have been conducted on the various revisions (1930, 1936, 1938, 1948) of the LIPS (Levine, 1982). However, only those reliability and validity studies dealing with the 1948 version are reported here. A 1976 study of internal consistency obtained a coefficient of .97 using the KR20 formula. Test-retest reliability studies with a variety of clinical populations have yielded coefficients ranging from .63 to .92. The one exception to this range was a test-retest reliability coefficient of .36 obtained by deaf preschoolers.

Levine (1982) reported 12 validity studies comparing the LIPS to the Stanford-Binet Intelligence Scale. Concurrent validity coefficients ranged from .38 to .89 with a median correlation of .77. Thus, the two tests are moderately and positively correlated. Correlations between the LIPS and the Full Scale IQ of the Wechsler Intelligence Scale for Children (WISC) were reported to range from .55 to .86 with five of the six correlations at .77 or above.

Clinical population comparisons of the LIPS and the PPVT (the original version of the PPVT–R that yielded a score that was called an IQ score rather than an age-equivalent score and a standard score as does the PPVT–R) have yielded correlations ranging from .37 to .58 with a median correlation of .57. Thus, the obtained IQ's on the two tests, although moderately correlated, do not appear to be interchangeable. This indicates that the LIPS and the PPVT tap different abilities, despite the fact that neither test requires verbal responses (Levine, 1982).

The LIPS is often used with autistic children to measure their cognitive ability (e.g. Maltz, 1981; Shah & Holmes, 1985), or to match them with other children for experimental purposes (e.g., Baron-Cohen et al., 1985, 1986). Shah and Holmes (1985) compared the performance of autistic children on the WISC–R and the LIPS. A high positive correlation (.82) was found between the LIPS IQ and the WISC–R Performance IQ, and a *t* test revealed no significant difference between the mean IQ scores. The LIPS IQ correlated highly with the WISC–R Full Scale IQ (.74), but the mean IQ scores were significantly different. There was only a low positive correlation between the LIPS IQ and the WISC-R Verbal IQ (.29) and the scores were significantly different.

Shah and Holmes (1985) noted the limitations of the LIPS which are listed below.

1. The scale appears to be measuring mainly nonverbal skills and may not be a good indicator of general intellectual ability.
2. There are too few items at each age level, which may lead to an inaccurate estimate of mental age. Also, the item difficulty level is not constant.
3. The construction of the scale does not allow any comparative analysis of different skills.
4. Unlike the WISC–R, the Leiter uses a ratio IQ. (p. 202)

Shah and Holmes (1985) concluded that despite its limitations, the LIPS yields IQ scores that are directly comparable to IQ scores obtained on the Performance Scale of the WISC–R, and that the LIPS can be regarded as a good indicator of nonverbal intelligence. Shah and Holmes (1985) also felt that since the LIPS takes less than half the time to administer than a full WISC–R, it is a practical instrument for research purposes with autistic children.

The LIPS was also administered and scored according to its accompanying manual in the current study. A nonverbal mental age was obtained from this test.

Language Sample

In addition to the administration of the LIPS and the PPVT-R, a language sample was elicited from each subject, with the assistance of a speech pathologist, to determine a mean length of utterance (MLU) for each subject. An individual's MLU is the average length of a speaker's utterances in morphemes and is a predictor of children's complexity of language (Owens, 1984). Its determination suggests not only what a child is capable of expressing verbally, but also what a child may be capable of comprehending. For example, subjects having MLUs of about 3.0 should be able to comprehend utterances of at least that long.

However, the longer the investigator's utterances are beyond 3.0, a greater likelihood exists that the subjects would have difficulty with comprehension. Thus, for the present study, determination of each subject's MLU assisted the investigator in assessing the relative difficulty of the tasks for each subject. The investigator was then better able to determine whether the linguistic requirements of the tasks may have been beyond a particular subject's comprehension level regardless of the A-R distinction concept.

The sample was obtained by having the subjects interact with the speech pathologist through playing with toy kitchen utensils and miniature imitation food, a number of knitted finger puppets representing various animals, and by discussing the activities portrayed in three large picture books. Each subject's MLU was obtained by dividing the number of morphemes uttered by the subjects, by their total number of utterances (or the first 100 utterances of more talkative subjects). Brown's (1973) stages of language development were then used to interpret the MLU information.

Scoring

For each appearance–reality task, the subjects had to correctly answer the pair of A-R questions to score a point. Therefore, the children could potentially score a maximum of three points if all three A-R pretest tasks (Size, Color and Object Identity) were answered correctly. Only children who scored zero were included as subjects.

Instructional Materials and Procedure

Materials

In accordance with Direct Instruction procedures, the materials and tasks selected to teach Color and Size appearance–reality distinction represented a range of positive examples of the concept taught. Each of the examples within each type of A-R distinction task was selected from different parts of the range. They utilized different objects in terms of appearance, color, and dimensions, but still maintained their status as positive examples of how sometimes things can look one way, but really be something else.

Three positive examples were selected for instruction from the range of each task. For the Color tasks, the instructional materials included the following: example A: a pink eraser, which, when placed under a blue filter appeared blue; example B: a yellow, plastic, two dimensional koala bear cutout (approximately 5.5 cm x 4 cm) that appeared orange when placed in a plastic spaghetti container containing a red filter; and example C: a white piece of paper (8 cm x 3 cm) that appeared green when placed behind a green filter.

For the Size task, the three examples used the following materials: example A: two stickers of dalmation dogs sitting on their haunches, one larger than the other (approximately 3 cm x 2.5 cm, and 2 cm x 2 cm, respectively); example B: two washers (approximately 3.5 cm, and 2.5 cm in diameter) affixed to yellow poster board; and example C: two sink plugs (approximately 5 cm, and 4 cm in diameter) also attached to yellow poster board .

In addition, to control for possible perseveration effects (i.e., subjects consistently pointing to the object on one side only) and thus to better determine skill mastery, examples B and C of the Size tasks each had two examples. Specifically, in one example the larger plug and washer were on the right side of their respective cards, while in the other example the larger plug and washer were on the left side. Both examples of the plug and washer tasks were randomly distributed during the instructional and testing sessions. All of the size tasks used the plexiglass magnifying apparatus.

In addition to the instructional tasks, two first order generalization tests within the range of examples (i.e., tasks using materials different from, but similar to the instructional tasks and demonstrating the same concepts as the instructional tasks), were also administered periodically throughout instruction. Materials for the first Color Generalization task (CG1) employed a piece of white chalk that appeared blue when a blue filter was placed over it, while materials for the second Color generalization task (CG2) included a purple crayon that appeared black, or dark green when a green filter was placed over it.

The first Size generalization task (SG1) used two rectangular pencil erasers of different sizes (approximately 4.5 cm x 1.7 cm, and 6 cm x 2 cm), also mounted on poster board. The second Size generalization task (SG2) used two "S" rings (metal in the shape of an S), the larger one measuring approximately 5.5 cm in length, and the smaller one measuring approximately 4.5 cm in length.

Figures 3.2 and 3.3 present the Color and Size tasks, and their respective generalization tasks in schematic form. The administration of the three examples of each task type, as well as the first order generalization tasks, comprised the baseline and intervention data points.

Procedure

Instruction took place in two, daily half-hour sessions, five days a week, until a criterion of three consecutive correct Color or Size A-R tasks (on tests of each instructional example) was reached, or 10 instructional sessions were completed. It was believed that this distributed practice would result in better retention of the concept than if instructional sessions were fewer and of longer duration (Becker & Carnine, 1980; Smith, 1978). Instruction continued on the task in the other domain until the criterion was reached on that one as well, or until 10 sessions were completed. Once criterion was reached on each of the

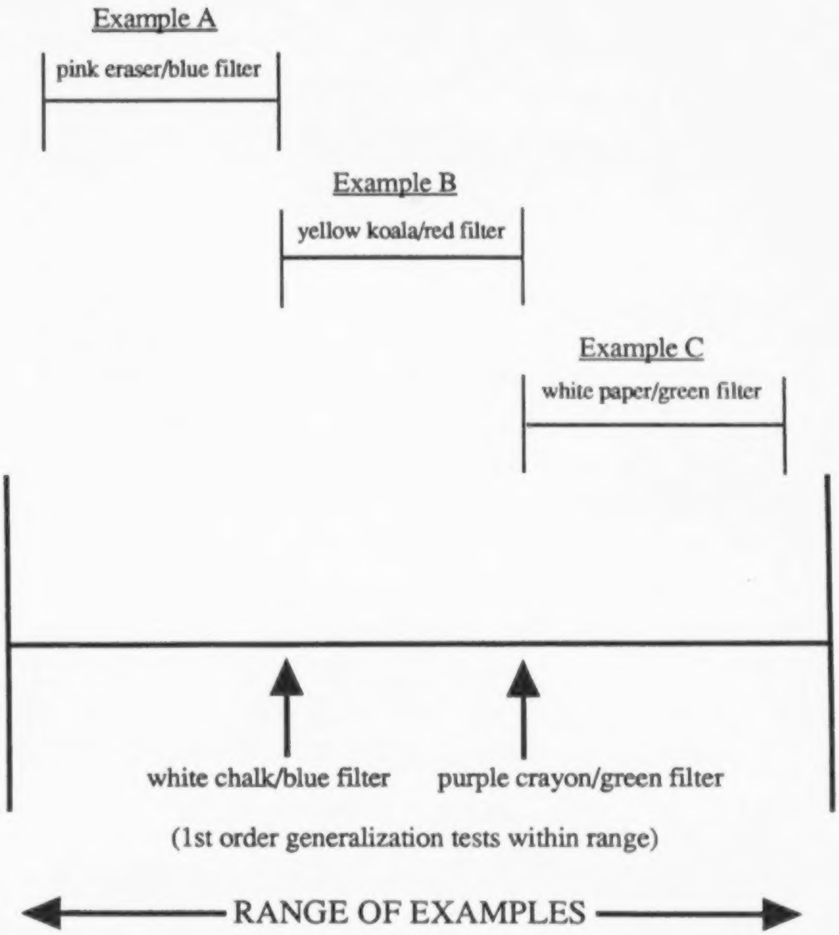

Example A
pink eraser/blue filter
Example B
yellow koala/red filter
Example C
white paper/green filter
white chalk/blue filter
purple crayon/green filter
(1st order generalization tests within range)
RANGE OF EXAMPLES

Figure 3.2. Color instructional tasks.

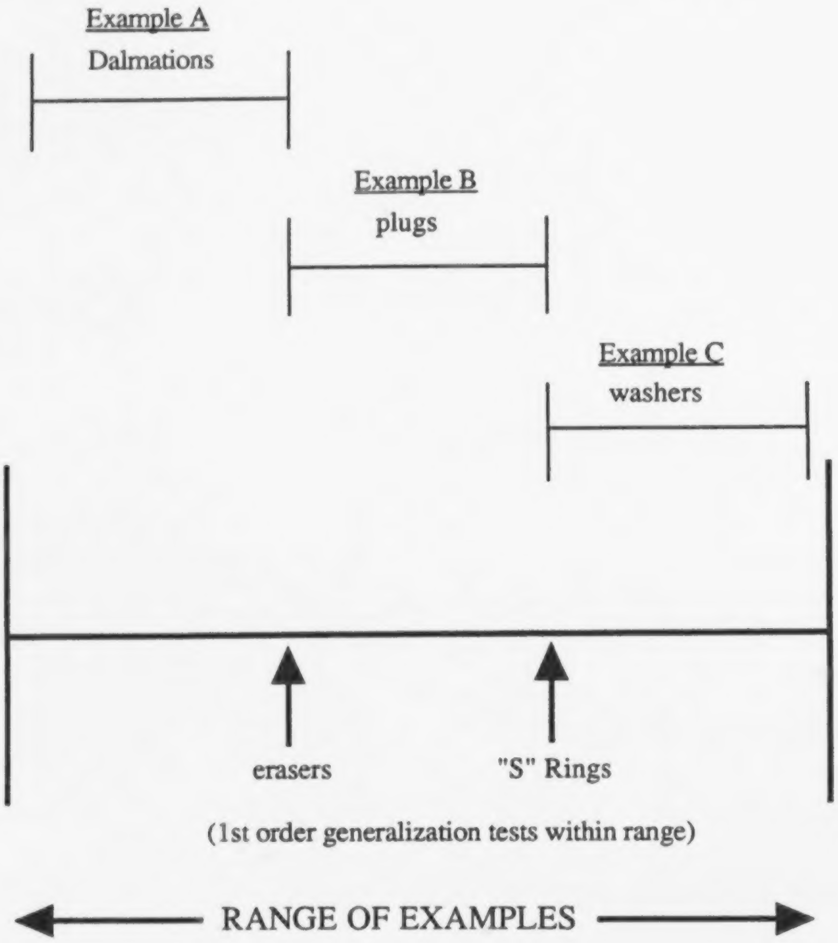

Example A
Dalmations
Example B
plugs
Example C
washers
erasers
"S" Rings
(1st order generalization tests within range)
RANGE OF EXAMPLES

Figure 3.3. Size instructional tasks.

tasks, a probe was taken eight days after the completion of instruction to check for maintenance.

Instruction within each session focused on only one type of task (either Color or Size). This method allowed the presentation of, and instruction in, two examples from within the range during each session, and served to reinforce the concept. The investigator believed that generalization was enhanced by emphasis on the fact that different tasks using different materials could follow the same A-R rule. The pairs of examples were presented in a random order from session to session. Total instructional time during each session was approximately 15 minutes. The remainder of the half-hour period was used to test the acquisition of the task (administration of examples A, B, and C in a test format), or the first order generalization tasks (CG1, CG2, SG1, or SG2).

Each instructional session was conducted as follows. The child was escorted into the room and sat on a chair across the table from the investigator. The child was presented with the first task (either a Color or Size example) and instruction began. Instruction on one example lasted approximately 7 minutes, at which time the second instructional example was introduced.

For the Color tasks, the general procedure was to present the tasks and require successively more participation on the part of the subjects in each subsequent phase. For example, the first presentation sequence was modelled completely by the investigator. In the second phase, the investigator provided the subjects with the answer at each stage of the presentation sequence, but then immediately asked the subjects what the correct answer was. Time delay, in which the investigator allowed successively longer periods of time to elapse before providing the subjects with the correct answer (e.g., 1-5 seconds), was used at this point. In the third phase, the investigator asked the subjects the appropriate question at each stage of the presentation, again employing the time delay procedure. Finally, in Phase 4, the investigator held the filter toward the subjects and asked them how they could make the object concerned “look” a different color. After this was completed correctly, the investigator then enquired as to the “real” color of the object. If the subjects answered this correctly, the investigator then asked them how they could show her the real color of the object. The subjects were expected to remove the filter to reveal the object’s real color.

An example of the script followed for the Color tasks is found in Appendix C, while Appendix D contains a task analysis sample of one instructional phase. If the subject made an error at any point in one phase, the experimenter reverted to the previous step that the subject answered correctly, and continued from there. If, at the end of the phase, the subject was unable to correctly answer the questions, the same phase was repeated before continuing to the next one.

Instruction for the Size tasks proceeded slightly differently. For example, in the modelling phase of these tasks, the card with the two mounted objects was presented on the bookstand. The subject was asked to point to which object was bigger (size discrimination ability). The plexiglass apparatus was then placed over the card so that the smaller object appeared larger than the objectively larger object. The experimenter pointed to the object under the magnifying glass (the phenomenally larger one) and said (with particular emphasis on the words in italics), "Here, this one *looks* bigger." The plexiglass was then removed and the experimenter pointed to the objectively larger object saying, "But this one is *really* bigger."

The original procedure that was to be used during the instructional periods, required the investigator to demonstrate the fact that one object was bigger than the other. The procedure entailed having the investigator, after having removed the plexiglass, place a piece of clear acetate over the objects. The acetate had two lines drawn on it, one line that corresponded to the common lower edge of the objects, and one line that corresponded to the upper edge of the larger object when it was placed over the objects. The investigator was then to draw the subject's attention to the difference in size between the objects by pointing out that whereas the upper edge of the larger object was flush with the upper line on the acetate, there was a space between the upper edge of the smaller object and the upper line. This procedure was attempted for the first few size instructional sessions with most of the subjects , but was abandoned for a number of reasons. First, the procedure was very awkward. It required the manipulation of a number of materials, which resulted in a substantial time delay between the investigator telling the subject that the phenomenally smaller object (when the plexiglass was in place), was *actually* the objectively larger object, and demonstration of the same. The possibility exists that this time delay prevented the subjects from understanding the relationship between the investigator's statement and its proof.

Secondly, such "proof" was considered unnecessary since, for inclusion in the study, the subjects had already demonstrated their understanding of the concept "bigger." Thus, it could be assumed that once the plexiglass was removed, the subjects would be able to correctly identify the larger object.

Finally, although the subjects were able to imitate the investigator in tracing the two lines on the acetate with their fingers, the investigator was not convinced that they understood why they were tracing the lines, since they could clearly see which of the objects was larger by looking at them. All of these reasons contributed to the elimination of the use of the acetate to "prove" the inherent "biggerness" of the larger object. Simply looking at the objects without the plexiglass was "proof" enough.

Similar to the Color instructional sequence, the experimenter modelled the first phase, and subsequently required more subject participation in determining which object was really bigger. A script of the Size tasks is also presented in Appendix C.

"Bigger" Tasks

As instruction proceeded, it became apparent that David, Brian, Bruce, and Steve may not have possessed a firm and consistent concept of the term "bigger." Thus, it became necessary to institute tasks that tested the subjects' understanding of the concept "bigger." Several pairs of objects, identical in all aspects except size, were used in these tasks. These objects included a pair of batteries (Sizes AA and D respectively), a pair of paperclips, a pair of file cards, and a pair of writing pads. Each task proceeded as follows. The investigator selected one pair of objects and placed them before the subject and the subject was asked to point to the object that was bigger. On subsequent trials each pair of objects was moved around on the table a number of times, with the larger object being randomly located to the left or right, above or below the smaller object. The subject was consistently asked to point to the larger object. This procedure allowed the investigator to determine if the subjects were able to point to the larger object consistently, thus demonstrating their understanding of the concept "bigger." This test also assisted the investigator to determine the validity of each subject's responses during the Size appearance-reality tasks. The "Bigger" task was administered as necessary throughout instruction, whenever the validity of a subject's responses was suspect.

Scoring

Scoring for the teaching tasks (when administered as tests) was similar to that of the pretests. That is, for each appearance–reality task, a subject had to correctly answer the pair of A-R questions to score a point. Scoring forms for the Color and Size tasks are found in Appendix B.

Because the Color and Size tasks for both the pretests and the teaching tasks required choosing between only two alternatives, it was possible for the subjects to be correct 50% of the time simply by chance. Therefore, it was necessary to control for chance responding. This control was accomplished by random repetition of the initial control question ("What color is this?" and "Point to the one that is bigger." or "Which one is bigger?" for the Color and Size tasks respectively), the Appearance question ("What color does it *look?*" and "Point to the one that *looks* bigger."), or the Reality question ("What color is it *really?*" and "Point to the one that's *really* bigger."). A particular task was considered unscorable when, upon question repetition, subjects appeared to be perseverating (i.e., answering the Size task questions by pointing only to the object on the left regardless of the question), or alternating their responses regardless of the question asked. The task was considered scorable when subjects' answers were consistent when the same question was being asked.

Experimental Design

The primary research approach used in the current study was a pre-experimental, case study. This design was selected for a number of reasons. First, case studies have the capability of providing knowledge about treatment effects that approaches the knowledge derived by true experiments (Kazdin, 1981, 1982). Whether or not such information is obtained depends on a number of variables including the presence of objective data, continuous assessment of the target behavior over time, determination of the stability of a particular behavior, the presence of immediate and marked effects once treatment is introduced, and the use of multiple cases (Kazdin, 1982). These design variables assist in ruling out such threats to internal validity as history, maturation, testing, instrumentation and statistical regression. The current study addressed each of these variables in its design as much as possible.

A second reason why a pre-experimental case study design was employed concerned the availability of subjects. Since autism is a relatively rare disorder (American Psychiatric Association, 1987), the probability of finding a group of autistic children that met the inclusion criteria was fairly low. This low probability, along with time constraints, made the feasibility of a group design difficult.

Finally, all previous theory of mind studies have used a group design in which group means were compared. It was this investigator's belief that a more thorough analysis of individual subject performance was required. For example, in a study examining autistic children's understanding of seeing, knowing, and believing, Leslie and Frith (1988) found that eight autistic subjects passed a limited knowledge task and five autistic children passed a false belief task. However, there is no indication of whether or not the five subjects passing the false belief task were the same subjects as those passing the limited knowledge task. Although a pre-experimental case study design per se does not specifically address this problem, the design allows the experimenter to determine which children pass which tasks, a determination that is not possible when only group means are compared. As well, while statistical analyses reflect the performance of a group as a whole (means and variances), they do not reflect the performance of any single subject. Analysis of group means alone does not allow the experimenter to determine how the intervention affects individuals, or the extent to which group performance represents individual subjects (Barlow & Hersen, 1984). A pre-experimental case study approach, on the other hand, has the capability of providing a more detailed analysis of individual performance (within subject comparison), in addition to a comparison of one subject's performance with another subject's performance (between subject comparison).

Elements of a multiple baseline design were also incorporated into the experimental design. Specifically, a baseline on both the Color and Size appearance-reality tasks was established for each subject, and intervention for most of the subjects was initiated in a sequential fashion. However, since intervention for subsequent subjects was not contingent on the previous subject reaching criterion (a requirement for a multiple baseline across subjects design), the current research cannot be considered a true multiple baseline design. Time constraints prevented the use of protracted baseline and intervention periods for the subjects.

Data Collection

Data points consisted of the assessment of the Color and Size task examples, and the Color and Size generalization tests. Initial baseline data for all subjects was collected on a daily basis, with the Color and Size tasks generally being administered one day, and the generalization tasks for Color and Size being administered the next day. This data collection continued until at least three data points for both the teaching tasks and the generalization tasks were collected, at which point intervention was initiated (i.e., A-B design).

In the course of the intervention period, it became evident that Nicky, Bruce, and Steve did not perform any differently than they did during the baseline period in one or other of the task domains. Thus, for these subjects, contingent secondary reinforcement was introduced into the Size and/or Color instructional sessions. Contingent reinforcement began with the introduction of instruction on the Size tasks for Brian and Steve. The addition of a reinforcement procedure made the study an A-B-C design. Specifically, tokens were awarded to the students contingent on attention and correct answers. The tokens were redeemable for a variety of edible treats (e.g., peanuts, raisins, popcorn) at the conclusion of the instructional session.

Once intervention began for each subject, the appearance-reality probe tasks continued to be administered approximately twice per week at the beginning of the instructional session, until a criterion of three out of three trials (i.e., tasks) was achieved on both the Color and Size tasks, and a criterion of two out of two correct responses were made on both the Color and Size generalization tasks, or until 10 instructional sessions on each of the Color task and the Size task had been completed. Once the criterion was achieved, intervention for that subject was discontinued and the investigator did not interact with the subject for eight days. After the eight day hiatus the subjects were again tested on the teaching and generalization tasks to test for maintenance of learning.

Alternate Question Format

As discussed in Chapter Two, Baron-Cohen (1989a), Flavell et al. (1983), and Flavell et al. (1986) believed that an additional control question (e.g., "When I take this away what color will it be?") was necessary in A-R tasks to ensure that the subjects understood that the true color of the object had not changed. These investigators considered this ability a pre-requisite to understanding of the appearance-reality distinction concept. The current investigator challenged this assumption, and explored its possibility through the random substitution of this alternate question format for the Reality question during administration of the Color and Size task tests.

Videotaping

Each subject was videotaped during all interactions with the investigator. While the videotaped instructional sessions were used solely by the experimenter to observe subject responses and behavior, the videotaped probe sessions were used to establish inter-rater reliability. Two videotaped probes per subject (Color and Size pretests) were used to train a second observer in the inter-rater reliability procedure described below. It was believed that these sessions sensitized the second rater to the subjects' idiosyncratic behaviors and speech, and provided sufficient practice to establish an acceptable degree of reliability between the raters. The remaining videotaped probes were used to calculate inter-rater reliability for each subject.

Inter-rater Reliability

The establishment of inter-rater reliability is important to minimize measurement error on the part of the experimenter. High inter-rater correlations indicate that the scores assigned to the subjects are not subjective and idiosyncratic to one rater (Smith & Glass, 1987), and increase the confidence that the observers recorded the same behavior at the same time (Tawney & Gast, 1984).

The second observer in this study was required to view videotapes onto which all testing of the subjects had been dubbed. While watching the videotapes, the observer scored each subject's answers using the same recording form used by the investigator during the study. A total of 494 scorable responses across all subjects and tasks were used in the calculation of inter–rater reliability. Ten additional probes were discarded because of inaudible replies on the videotape, missing data from the second observer, or a camera angle that prevented the second observer's proper viewing of a subject's responses.

The second observer was trained in the scoring procedure by viewing each subject's Color and Size pretest responses, and with the aid of the investigator, scored each response. In addition, the investigator provided the second rater with other examples of scoring that would be encountered in the course of viewing. As evidenced by the results of the inter–rater reliability measures, this training was adequate.

In this study, the point–by–point method was used to determine total percentage of agreement, effective percentage of agreement on occurrences (in this case the correct answer to the appearance–reality questions), and effective percentage agreement on nonoccurrences (incorrect answers to either the appearance, reality or both questions). These calculations were carried out on all teaching and generalization tasks for each subject. As Hartmann (1977) noted, calculating effective percentage agreements provides a more sensitive measure of observer reliability by excluding agreements on occurrence and nonoccurrence that may be due to chance. In addition, both phi and kappa coefficients were calculated. Phi is the product–moment correlation between two sets of dichotomous (occurrence–nonoccurrence, yes–no) data, while kappa indicates the proportion of agreements corrected for chance agreements (Cohen, 1960; Hartmann, 1977). In situations where the rate of occurrence of the target behavior is almost identical between the two observers, these statistics are nearly identical in value, and can be used interchangeably. In situations where phi equals kappa, phi can also be interpreted as a corrected percentage agreement statistic (Hartmann, 1977).

The results of the inter-rater reliability measures are presented in Chapter Four as each subject's performance is discussed.

Ethical Considerations

Parents of the subjects were fully informed as to the nature and methodology of the study, its risks and benefits, and completed a consent form. Both the information provided to the parents and the consent form are presented in Appendix E. Parents were free to withdraw their child from the study at any time. Permission to conduct the study was also obtained from the Edmonton Public School Board and the Ethics Committee of the Department of Educational Psychology at the University of Alberta. The children's permission was also obtained through verbal request.

Prior to signing the consent forms, the parents were advised of the potential risks and benefits involved in their child's participation in the study. Since all the testing and instruction in this study was conducted in a private room, it necessitated the children being taken from their regular classroom two half-hour periods a day, five days a week, for a maximum of 20 sessions. The respective teachers were consulted as to when would be the most convenient time to conduct the sessions. Regardless of the time scheduled for instruction, the children necessarily missed some part of their classroom routine. However, the researcher believed that the benefit of individual attention and instruction outweighed the disadvantages of removing the children from their classrooms. If instruction was successful, the children would have learned important skills that serve as a basis for conceptual perspective-taking. If the theory of mind instruction was unsuccessful, however, the children would still have benefitted from additional one-on-one instructional time emphasizing attention, eye contact, and verbal responses.

Because this study required the participation of children diagnosed as autistic, it was necessary to gain access to confidential school records concerning the children's diagnoses. The appropriate release form is also presented in Appendix E.

As discussed above, all of the sessions with each child were videotaped. However, these tapes remained confidential and were viewed only by the researcher, the inter-rater observer, and the researcher's committee. Anonymity of the subjects has been maintained through the use of pseudonyms.

CHAPTER FOUR

RESULTS

This chapter presents the results of the study. In the first section of the chapter, the data for each subject are presented and considered. Presented first is a description of each subject in terms of diagnostic and background information, and a description of the subject's behavior during the study. Next, the results of the inter-rater reliability calculations are discussed. Each subject's performance in the two task domains targeted for instruction is presented and discussed in the third section. In addition, each subject's performance on the Object Identity tasks is discussed, along with the observed relationships between the subject's performance on the various tasks (intrasubject comparison). The chapter concludes with a summary of the obtained results.

Individual Subject Data

Before communicating the study's results, it is necessary to discuss the information that is provided on the graphs to be presented. Figure 4.1 on page 69, will assist readers in understanding the following description of the graphs' features. The asterisks located below particular session numbers on each of the Color and Size task graphs indicate days on which no subjects were seen because of professional development days, school holidays, or the investigator's inability to see subjects on that day. The short vertical lines between session numbers indicate the end of a school week, and thus, the point at which the subjects were not seen for two days. The dates on which the subjects were seen by the investigator, and their corresponding session numbers, are presented in Appendix F.

The numbers located on the ordinate of the Color and Size graphs for each subject represent the number of tasks on which the subjects gave correct answers on either the teaching tasks or the generalization tasks when either the standard format (Appearance and Reality questions asked for each task), or the alternate format (Appearance question asked, followed by "If I take this away what color will it be?" ["If I take this away which one will be bigger?"]) was used. In each task domain, there was a total of 3 teaching task examples, with each task worth a possible score of 1 (correct answers on both the Appearance and Reality questions of a task earned 1 point). Therefore, the maximum possible score for the teaching tasks was 3. Each generalization task was also worth 1 point. Thus, the maximum possible score on the generalization tasks was 2. The subjects' responses to the alternate question format are presented on their respective graphs to provide readers with an indication of the effectiveness of this format on the subjects' ability to complete the A-R tasks correctly, compared to the effectiveness of the standard question format. Because the alternate format was not used an equal number of times with each task or subject, nor, in most cases, was it employed during the intervention period, conclusions cannot be reached regarding the relative effectiveness of the alternate format, or its necessity of inclusion as a control question as was suggested by Baron-Cohen (1989a), Flavell et al. (1983), and Flavell et al. (1986). Because the investigator's primary interest was in comparing the present subjects' performance with that of subjects in previously published studies (e.g. Baron-Cohen, 1989a, Flavell et al., 1986), only probes using the alternate question format were attempted. However, the data that was collected using this format can be used as evidence for the necessity of the alternate question format's more systematic inclusion in future research.

Because of the difficulty in determining how many tasks were completed using either the standard or alternate question format when both were used in a given session, accompanying tables provide specific information on the subjects' performance during these sessions.

The Object Identity generalization task (the Rock task) information is presented on both the Color and the Size task graphs of those subjects to whom the task was presented. This presentation format allows the reader to easily determine the subjects' performance on this task relative to their performance on the Color and the Size tasks. The Rock task was administered only to those subjects who achieved the criterion (3 out of 3 tasks) on either the Color or the Size tasks; the task had a maximum possible score of 1.

Upon examination of the subjects' Color and Size task graphs, the reader will notice that various sessions are labeled "unscorable." Responses were considered unscorable if the subjects perseverated (e.g., pointed only to the object on the left side on the Size task regardless of the question being asked), or if the subjects simply alternated responses (e.g., consistently picked the alternate choice even when the same question was asked repeatedly). In some testing sessions, only one task might have been unscorable. In this situation a cumulative task score was assigned and graphed, the unscorable task thus being undetectable upon examination of the graph. These instances are explained as necessary.

In addition to the graphs depicting the subjects' performance during testing of the Color and Size tasks, graphs illustrating the subjects' performance during Phase 4 of the intervention sessions are also presented (refer to Figure 4.2 for an example of a Phase 4 intervention graph), to permit a more detailed look at each subject's performance on the specific tasks within the domain concerned. As explained in the previous chapter, Phase 4 was essentially a testing phase, presented after instruction on the task had taken place. To reiterate, this phase required the subjects to place the colored filter (or the plexiglass) over the object to make it look a different color (or larger). The subjects were then required to tell the investigator what they had done, answer a Reality question, and demonstrate how they knew the object's true color (or size) by removing the filter (or plexiglass). To achieve a score of 1, it was necessary for a subject to complete each of these steps correctly on a given task. Two task examples were presented in any given instructional session, and the subjects' performance on each of these tasks is reflected on the Phase 4 intervention graphs. The session numbers that are identified on the abcissas of these graphs reflect the intervention session number only for the particular subject concerned. Hence, Session number 1 for one subject was not necessarily Session number 1 for another subject. However, once intervention was begun with a particular subject, each session was numbered sequentially from that point, regardless of whether the subject worked with the investigator on a given day.

For those subjects with whom the reinforcement procedure was implemented, the session in which reinforcement began is indicated by a solid vertical line on both the Color and Size graphs, as well as the Phase 4 intervention graphs. The presence of this demarcation should assist in the interpretation of the effectiveness of the reinforcement procedure for the individual subjects.

Nicky

Description

With a chronological age of 9.4 at the beginning of the study, Nicky was the youngest subject to be included. Nicky was first diagnosed at the age of 14 months by a local hospital as having cerebral hypotonia with a global developmental delay. The delay was associated with congenital hydrocephalus and cerebral atrophy. Early intervention for Nicky began at 14 months of age, and in 1983 she was enrolled in an early education program at a public school, a program in which she remained for 3 years. In 1985 (at age 3), upon administration of the autism checklist, the local public school board determined that Nicky qualified for services in the category of severe behavior disorder—autism.

Autistic behaviors have been observed in Nicky since infancy, and she has engaged in self-abusive behavior since the age of 18 months. This behavior consists of cracking her nose against her knee, or on a table, and of banging her head and chin against hard surfaces. This behavior reached a peak in 1988, with a recorded frequency of 4800 times per hour. Although Nicky used to wear a helmet for self-protection, her self-abusive behavior had decreased so dramatically in her current placement this year (only a few times per month), that a helmet was no longer considered necessary. As well as being on a behavior modification program, Nicky is on medication (fluphenazine) to reduce self-abusive behavior. Information on the present frequency of Nicky's self-abusive behavior was not available to the investigator. When Nicky's self-abusive behavior became too difficult for her mother to handle, Nicky was placed in a group home where she now resides. The year of admission to the group home is not known to the investigator. Nicky's mother, however, remains very interested, and involved in Nicky's program. Nicky is currently enrolled in a program for autistic students in a regular public school, and is integrated into a regular grade one class for mathematics.

Intellectual testing conducted in 1987 found that with a chronological age of 5.2, Nicky's mental age as measured by the LIPS was 5.3, and her receptive vocabulary age-equivalent on the PPVT-R was 3.3. Nicky's expressive language age-equivalent, as measured by the Merrill Palmer, was between 2 and 2.5 years.

On the inclusion criteria assessment for the present study, with a chronological age of 9.4, Nicky achieved a mental age of 5.3 on the LIPS and a receptive vocabulary age-equivalent of 4.5 on the PPVT-R. It should be noted here that the investigator believed that higher mental ages and receptive vocabulary age-equivalents would have been obtained for all the subjects in the study if the test administration had not had to follow standardized formats. For example, there were numerous instances on the LIPS where the subjects seemed to make careless errors, either mixing up two blocks or placing a block in the slot upside down. On items at the higher age levels, these actions resulted in item failure, and consequently a lower mental age estimation. Had the investigator been able to direct the subjects' attention to each block configuration, the subjects may have been able to identify and correct some of these errors. Nicky's mean length of utterance (MLU) on a language sample obtained with the assistance of a speech pathologist was 4.17.

Nicky came willingly with the investigator for the first, and all subsequent sessions. She was very cooperative and conversational during the sessions, for example, asking the investigator about the room she was using, and what the investigator's position at the school was. Throughout the study, Nicky would make comments that were irrelevant to the task at hand. For example, during the administration of a PPVT-R item, Nicky suddenly said, "When you grow up you get to learn things." During another item she said, "I got nice hands." When such irrelevant remarks were made, the investigator prompted her to pay attention to the task at hand, and repeated the item.

Nicky demonstrated a lot of pronomial reversal in her speech but it was not constant, even within one utterance. For example, Nicky frequently asked about what was going to happen after we had finished working for the day: "When you're done, you go back to class. Going back to my class." Nicky also made many comments, in the course of the study, about the condition of her nose, which was very bruised and swollen from self-abuse: "Your nose gonna heal up pretty soon," and "Is your nose healing up? A little bit." And later in the study, "See, your nose is all better. Your nose is all better, right?" Nicky has an imaginary friend named Heldon who apparently needed a helmet, and presumably hit himself: "The helmet's for Heldon. The helmet. He's got that on. He's bad."

Nicky seemed to need constant reassurance regarding the sequence of activities. This behavior continued for the duration of the study. For example, Nicky would say, "After we do some stuff we go back to the room. Pretty soon. It's not long. After this, right?" The investigator informed Nicky about what was going to happen as specifically as possible.

Throughout the study Nicky displayed some self-stimulatory behavior. This took the form of flicking at her hair with her fingers, rubbing her hands together vigorously and jerking her head forward, almost in the manner of a petit mal seizure. During the period in which she was jerking her head, Nicky seemed to be unaware of her surroundings, which necessitated the repetition of questions and tasks.

On occasion, during administration of the Size tasks, Nicky would point, or begin to point, to one of the alternative answers before a question had even been asked. This behavior was dealt with by a verbal prompt to "wait until you hear the question," and to "listen to the question first."

Inter-Rater Reliability

Color tasks. Table 4.1 presents the inter-rater reliability data of the Color tasks from which percentages and coefficients were calculated. The two-by-two table presented in this figure summarizes the two kinds of agreements and two kinds of disagreements that are possible when scoring the occurrence and nonoccurrence of a single behavior. As explained in the previous chapter, correct responses on both the Appearance and the Reality questions was considered an occurrence of behavior (1), while incorrect responses on one or both of the questions constituted a behavior nonoccurrence (0).

In Table 4.1, and in all subsequent references to inter-rater reliability scores, Observer 1 was the investigator, while Observer 2 was the second rater who scored the responses from the subjects' videotapes. The numbers above and beside the table (0 and 1) refer to the rating the observer gave to the behavior, either an occurrence of the behavior (1), or a nonoccurrence of the behavior (0). The numbers within each cell represent the frequency with which each type of agreement or disagreement was observed. For example, the frequency in Cell B (23) indicates the number of occasions for which both observers recorded an occurrence of the behavior, while the number in Cell C (30) indicates the number of occasions for which both observers recorded a nonoccurrence of the behavior. The remaining cells, (Cells A and D) indicate the frequency with which disagreements between the observers were recorded. The marginal values (e.g., A + B) indicate the sums of the rows and columns of the table necessary for the calculation of percentages and coefficients.

Table 4.1

Nicky—Inter-rater Reliability on Color Tasks

		Observer 2		
		0	1	
Observer 1	1	A 1	B 23	A + B = 24
	0	C 30	D 0	C + D = 30
		A + C = 31	B + D = 23	N = 54

As seen in Table 4.1, out of total of 54 observations of Nicky's performance on the Color tasks, there was one disagreement in which the investigator (Observer 1) recorded an occurrence of behavior, while the second observer recorded a nonoccurrence. There were 23 agreements of behavior occurrence, and 30 agreements of behavior nonoccurrence.

Total percent agreement (the percent of total observations the observers agreed) on the Color tasks was 98, effective percent agreement on occurrence (the percentage of intervals in which both observers agreed that the behavior occurred) was 95.83, and effective percent agreement on nonoccurrence (the percentage of intervals in which both observers agreed that the behavior did not occur) was 96.77. Phi (the product–moment correlation between two sets of dichotomous data) and kappa (the proportion of agreements between the observers, corrected for chance agreements) correlations were also calculated for all subjects. The phi correlation coefficient for Nicky on the Color tasks was .96, and the kappa correlation coefficent was .98.

Size tasks. Inter-rater reliability data of the Size tasks for Nicky are presented in Table 4.2. Out of a total of 48 observations of the Size tasks, there were no disagreements where either the investigator recorded an occurrence of behavior and the second observer recorded a nonoccurrence, or vice versa. There were 23 occurrence agreements, and 25 agreements on behavior nonoccurrence. As a result of the complete agreement between the observers on the occurrence and nonoccurrence of behavior, total percent agreement, effective percent agreement on occurrence of behavior, and effective percent agreement on nonoccurrence of behavior were all 100. The resulting phi and kappa coefficients were 1.0.

Table 4.2

Nicky—Inter-rater Reliability on Size Tasks

		Observer 2		
		0	1	
Observer 1	1	A 0	B 23	A + B = 23
	0	C 25	D 0	C + D = 25
		A + C = 25	B + D = 23	N = 48

Results.

Color tasks. Throughout the baseline period on the Color tasks (Figure 4.1), Nicky consistently made phenomenist errors on each of the tasks when the standard question format for the Reality question ("What color is it *really*?")was used. That is, she reported that the apparent color of each object was also the true color of the object. However, when the investigator used the alternate question format ("If I take this away [indicating filter], what color will it be?"), Nicky often completed the task correctly. As Table 4.3 illustrates, of the 9 task trials on which the alternate question format was employed, Nicky correctly identified the true color of the object concerned on 5 of them. On the Eraser task in Session 4, Nicky first gave a phenomenist answer to the alternate question format (the eraser would be purple when the filter was removed) but then quickly changed her answer to "pink."

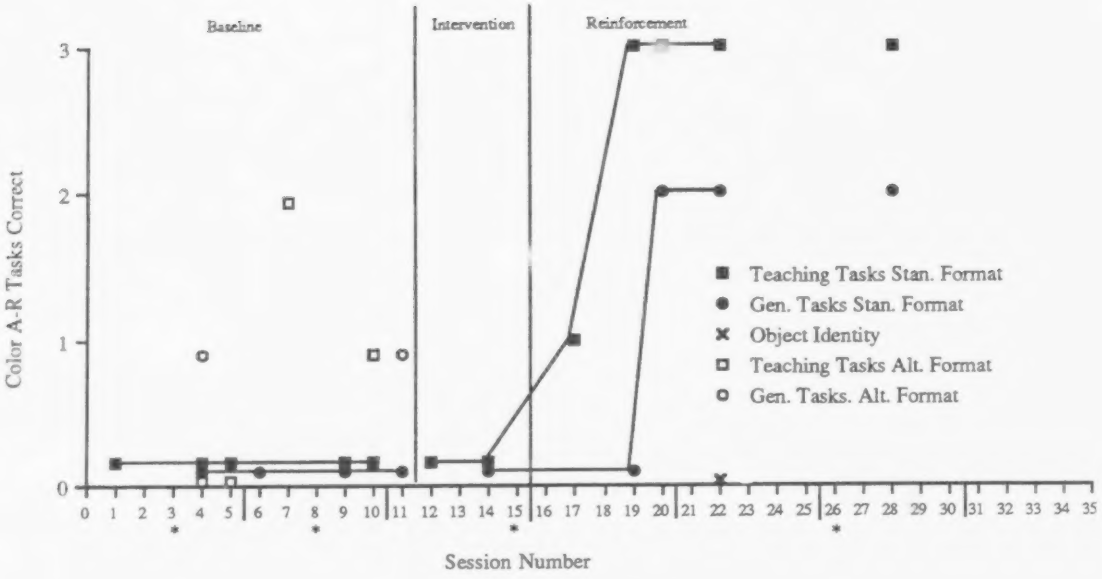


Figure 4.1. Nicky—Number of correct responses on tests of Color tasks across conditions.

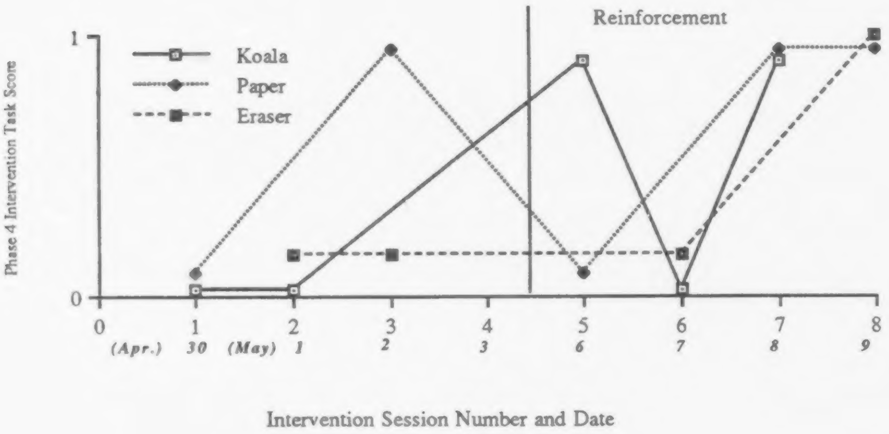


Figure 4.2. Nicky—Phase 4 intervention on Color tasks.

Table 4.3

Nicky—Data Breakdown of Color Sessions When Alternate Question Format Used

		Standard Format		Alternate Format	
Session Number	Task	Pass	Fail	Pass	Fail
4	Paper		√		
	Eraser				√
	Koala				√
	Chalk (CG)*		√		
	Crayon (CG)			√	
5	Eraser				√
	Paper		√		
	Koala		√		
7	Eraser			√	
	Koala				√
	Paper			√	
10	Eraser			√	
	Paper		√		
	Koala		√		
11	Chalk (CG)			√	
	Crayon (CG)		√		

*Color Generalization Task

Despite her spontaneous correction, the item was scored conservatively as incorrect, because it could not be ascertained whether Nicky really knew the correct answer to the question, or was simply choosing the alternative. However, Nicky's performance on the remainder of the tasks presented using the alternate format indicated that this format was somewhat easier for her to understand than was the standard format. Although her performance on the alternate format tasks did not exceed chance level (55%), it was significantly better than was her performance when the standard task format was used (0%). As seen in Figure 4.1, Nicky reached criterion (3 out of 3 teaching tasks correct) on the Color teaching tasks by Session 19, the 8th session after intervention had been introduced. Reinforcement, in the form of tokens that were exchangeable for snack food, was introduced in the 6th session after the introduction of intervention (essentially the 5th session since no subjects were seen in Session 15). Prior to this session, Nicky was unable to complete any of the tasks correctly. After secondary reinforcement was introduced, however, Nicky was able to answer the Appearance and Reality questions correctly for one of the tasks on the next testing session, and correctly answered the questions on all three Color teaching tasks in the testing session subsequent to that. By the 5th session after reinforcement had been introduced, Nicky answered the A–R questions correctly on the two generalization tasks as well. Nicky continued to achieve at ceiling on the Color teaching tasks and the Color generalization tasks for the remainder of the study.

Examination of Figure 4.2 reveals that at least two of the tasks (Koala and Paper) appeared to be of the same relative difficulty for Nicky to understand. During some of the instructional sessions she completed Phase 4 correctly on these tasks, but was unable to do them correctly during other sessions. The investigator believed that this rather erratic responding could be attributed to the nature of the task. In answering the Reality question, there was essentially only a choice between two colors, since for any given task only two colors had been mentioned. Thus, it was quite possible that by chance, Nicky would give the correct answer 50% of the time. It was not until a consistent pattern of correct responding was established, that one could conclude that Nicky truly understood the appearance–reality concept. For example, when Phase 4 of the Paper Color task was presented during the 3rd instructional session, Nicky, upon the verbal prompt, “Show me how you can make this look green,” took the filter and put it over the piece of paper. When the investigator asked, “What did you do Nicky?”, she replied, “Made it look green.” The investigator then asked, “But what color is it *really?*” (thus even implying that a different answer was required). “It’s *really* ...”, “Green,” replied Nicky. The investigator then said, “No...” and began to take the filter away. Nicky then quickly changed her answer to “white.” When Phase 4 was subsequently repeated, Nicky answered the Reality question correctly. So, although this task was scored as correct in Figure 4.2 (since the last presentation of this phase elicited a correct answer), the more likely explanation is that Nicky was simply selecting the alternative choice, rather than demonstrating understanding of the concept.

The only Color task that Nicky consistently got incorrect during Phase 4 was the Eraser task. The possibility exists therefore, that Nicky found this task more difficult than the other tasks (tasks which she was able to complete correctly at least some of the time), and that once she understood the Koala and Paper tasks, she was able to generalize her understanding of the appearance–reality concept to the Eraser task.

During the first Color intervention session, Nicky spontaneously covered her own hand with the green filter after the investigator had finished the Paper task, so the investigator proceeded with the task using Nicky's own hand (the true color of which, perhaps, was more familiar to her than was the color of the objects used). After Nicky put the filter over her hands, she stated, "green hands." The investigator asked the Appearance question, "What color do your hands *look*?", to which Nicky replied, "green." The investigator then said, "Right. But they're not *really* green. What color are your fingers *really*?" and Nicky answered, "peach." It is possible that Nicky was able to correctly answer the Reality question because the true color of the rest of her arm remained visible, and this visual cue provided her with enough information to answer the question correctly, especially since the color "peach" had never been mentioned before. However, at Nicky's request, the task was immediately repeated using a blue filter, and Nicky gave a phenomenist answer to the Reality question.

Nicky was quite interested in using the filters to change the color of her hands, her shirt, and a pencil that was on the table, and did this during three different instructional sessions. Although she ultimately gave a phenomenist response to the Reality question on these occasions, there were several instances in the fourth intervention session where Nicky, imitating the investigator, "taught" the Paper task to her imaginary friend, Heldon, and answered the Appearance and Reality questions correctly (although she never covered the paper completely with the filter). The following videotape transcript excerpt illustrates how Nicky used the paper and the green filter to teach Heldon.

> It sort of looks one way. I'm going to do this with Heldon. (Picks up filter). Heldon, show me the one that looks one way but it'll be something else Heldon. Show me. (Places filter over half of the paper. Points to green section). This looks green but it's really white (Points to white section). Really white. I'll take this away and it'll be something else but it'll look one way. This looks green. (Points to green section). It looks green Heldon, but (points to white section), this is really white Heldon.

Despite this apparent understanding of the task, it appeared that Nicky was simply imitating the procedure without comprehension of the concept because when Phase 4 was repeated (with the paper completely covered by the filter), Nicky said that the paper was "really" green. This interpretation is supported by the investigator's observations of other sessions in which Nicky either repeated the investigator's words or, after one presentation, said them with the investigator, but continued to answer the A–R questions incorrectly during Phase 4, and on the testing of the items. For example, during Phase 2 of the 4th instructional session, the following exchange took place during the presentation of the Koala Color task:

> I (Investigator): (Puts yellow koala in container with red filter so that about half of it looks red) It looks red. What color does it look?
>
> N (Nicky): It looks red.
>
> I: But really it's yellow. What color is it really?
>
> N: But really it's yellow.

Thus, although demonstrating syntactic competence with the linguistic form of the tasks, Nicky's semantic understanding of the tasks was questionable, given her performance on the items when they were administered in a test format.

It is possible, however, that this repetition procedure ultimately assisted Nicky in understanding the A–R concept, since by the 7th intervention session she was able to complete Phase 4 of all the Color tasks correctly (Figure 4.2). This mastery is reflected in Figure 4.1 as well, with Nicky performing at ceiling on the teaching tasks the next time they were tested (Session 19) and on the generalization tasks the subsequent day.

Another interesting behavior that Nicky exhibited was noted during Phase 3 of the instructional sessions, the phase in which Appearance and Reality questions were asked of the subjects without the correct answer being modelled. During one presentation of Phase 3 of the Eraser task in intervention Session 2, for example, the investigator gradually covered the pink eraser with the blue filter always asking the Reality question ("What color is it really?"), so that only one answer was required. Nicky answered the question correctly ("pink") until the eraser was completely covered by the filter, and then said that the real color of the eraser was blue. This behavior led the investigator to conclude that Nicky either did not have a functional understanding of the lexical term "really," or that she believed that the veridical color of the eraser changed from pink to blue once the filter was covering it completely.

In summary, although Nicky was initially unable to complete the Color tasks correctly, within 8 intervention sessions criterion had been reached on the teaching tasks. Nicky was also able to complete the generalization tasks correctly in the subsequent session (Session 20).

Size tasks. As seen in Figure 4.3, Nicky was unable to do the Size tasks during the baseline period and, similar to her performance on the Color tasks, consistently made phenomenist errors regardless of whether the Reality or the Appearance question was asked first, and regardless of the side of the display on which the objectively larger object was located. Of the five occasions during the baseline on which the alternative question format was used (Table 4.4), Nicky was able to complete the task correctly on only one occasion

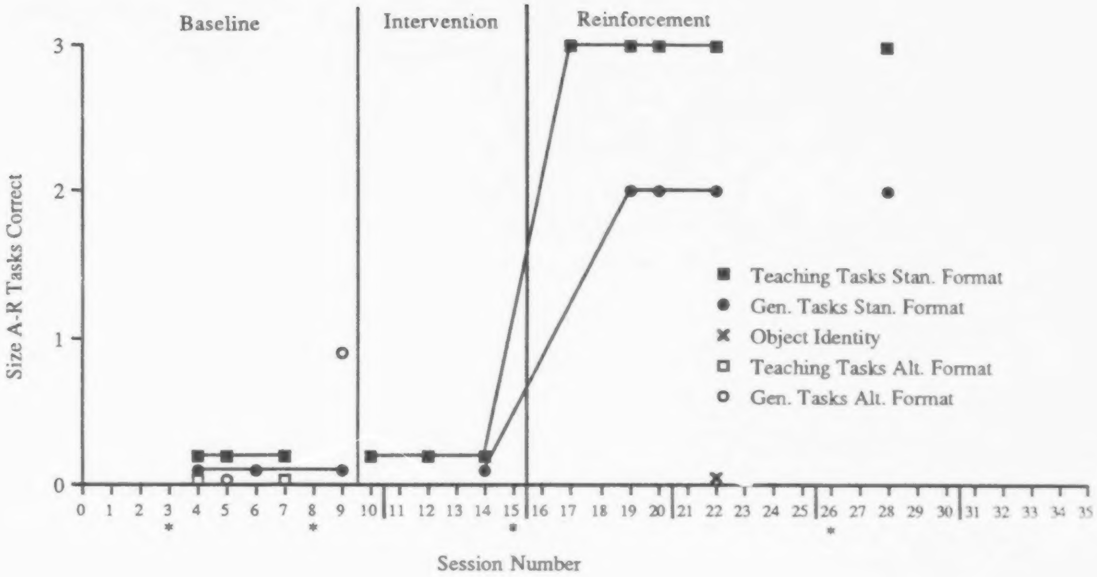


Figure 4.3. Nicky—Number of correct responses on tests of Size tasks across conditions.

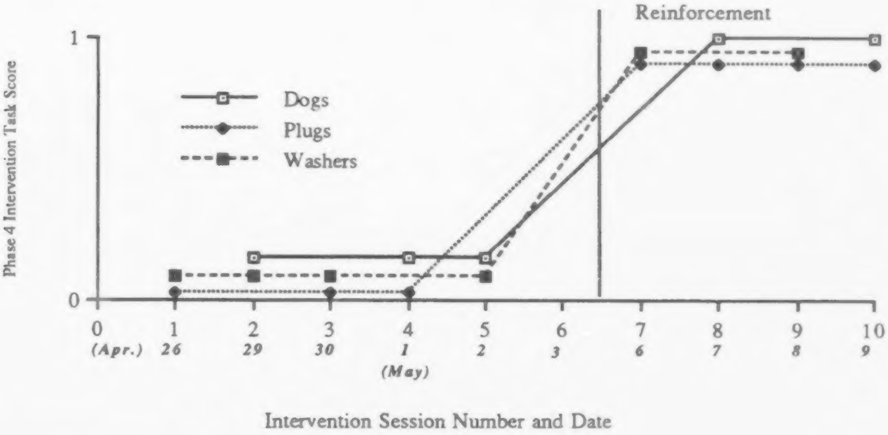


Figure 4.4. Nicky—Phase 4 intervention on Size tasks.

Table 4.4

Nicky—Data Breakdown of Size Sessions When Alternate Question Format Used

		Standard Format		Alternate Format	
Session Number	Task	Pass	Fail	Pass	Fail
4	Plugs		√		
	Washers				√
	S rings (SG)*		√		
	Erasers (SG)		√		
	Dogs				√
5	Washers		√		
	Dogs		√		
	Erasers (SG)				√
7	Plugs		√		
	Dogs				√
	Washers		√		
9	Erasers (SG)			√	
	S rings (SG)		√		
	Erasers (SG)		√		

*Size Generalization Task

(20% success rate). Thus, unlike her performance on the alternate format Color tasks, use of the alternate format on the Size tasks did not improve Nicky's performance in any appreciable way.

By Session 17, however, (Figure 4.3), Nicky had reached criterion on the Size teaching tasks. Mastery of the Size generalization tasks was demonstrated during their next assessment (Session 19).

Figure 4.4 illustrates that Nicky's understanding of the concept, as demonstrated in Phase 4, was not specific to any particular task or location of the larger object, but that once she could apply her understanding of the concept to one Size task, she was able to apply it to all Size tasks. Comparison of Figure 4.3 and Figure 4.4 suggests that Nicky would have been able to pass the Size tasks in Session 16 (May 16 in Figure 4.3), had they been administered, since consistent responding during Phase 4 of the intervention had been established at this point. However, because all of the Size tasks were not administered in test format until Session 19, this session is considered the one in which mastery of the all the Size tasks was achieved.

The investigator noted a number of interesting behaviors as instruction on the Size tasks proceeded with Nicky. The first was Nicky's insistence on pointing to the phenomenally larger object during Phases 2 and 3, even after the correct choice had been repeatedly modelled and "proven." For example, during Phase 2 of instructional Session 4, the investigator informed Nicky of the correct choice when an immediate response was not forthcoming. The investigator immediately asked the Reality question again, and Nicky still chose the phenomenally larger object. The investigator then removed the plexiglass, and asked Nicky which object was larger. Nicky pointed to the correct object. The question was repeated, and again, the correct response was elicited. The investigator replaced the plexiglass over the objects, and repeated the Reality question for the third time. This time, Nicky pointed to the phenomenally larger object rather than to the objectively larger object. On occasion, the investigator continued to point to the correct answer while she asked Nicky the Reality question and waited for her to respond. Despite this gestural prompt, Nicky would still choose the phenomenally larger object. As will be seen, this behavior was also evident in most of the other subjects as well.

As observed during instruction on the Color tasks, Nicky imitated the investigator's instructions in order to "teach" Heldon the Size tasks. For example, during Phase 4 of the first intervention session, when the washers were used, Nicky correctly placed the plexiglass over the objects in response to the prompt, "Use this (investigator indicates plexiglass) and show me how you can make this one (investigator points to smaller object) *look* bigger." In response to the question, "Now which one *looks* bigger?," Nicky replied, "Heldon, which one looks bigger. Point to the one that looks bigger." This response again showed that Nicky had no problem in repeating what was required of her, but it did not confirm that Nicky *understood* what was required of her. Interestingly, however, unlike her performance during intervention on the Color tasks, Nicky did not model the complete procedure for her imaginary friend, but only gave the instructions.

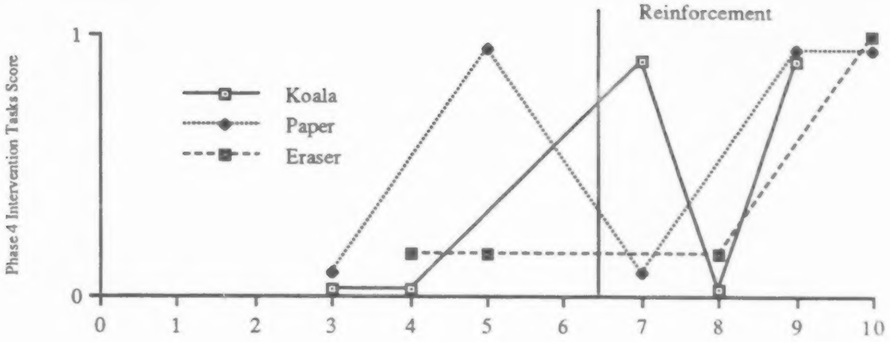
As in the Color tasks, Nicky also demonstrated a lot of indecision during Phase 4 of the Size tasks before demonstration of understanding of the concept was established. Nicky often alternated her choice of which object was "really bigger." For example, when the plugs were used during Phase 4 of intervention Session 4, Nicky, in answering the Reality question when the plexiglass was not covering the objects, pointed correctly to the objectively larger object. When the same question was asked immediately after the plexiglass was placed over the objects, Nicky pointed to the phenomenally larger plug, but then alternated, and pointed to the objectively larger plug. However, once understanding of the concept *was* established, Nicky answered both the Appearance and Reality questions consistently and accurately.

Evidence of Nicky's understanding of the concept was particularly strong in intervention Session 10 (Figure 4.4). During Phase 4 of this session, Nicky placed the plexiglass over the plugs as requested. However, on this occasion the plexiglass was upside down (the investigator had failed to orient the plexiglass correctly, as she usually did), so that when Nicky placed it over the plugs, the bigger plug was made to look even bigger than the smaller plug (as opposed to making the smaller plug appear larger than the objectively larger plug). The investigator proceeded with the task, however, to see how Nicky would respond. The investigator pointed to the objectively smaller plug (and, in this situation, the phenomenally smaller object as well), and asked, "Does this one *look* bigger?" Nicky appeared confused by the situation and did not respond. The investigator then said, "Point to the one that looks bigger," and Nicky correctly pointed to the phenomenally larger plug (and objectively larger plug). The plexiglass was subsequently oriented correctly, and the Appearance and Reality questions were asked, and answered correctly by Nicky. When Nicky was then asked to "Show me how you know it's bigger," she removed the plexiglass and pointed to the bigger plug. The investigator believes that Nicky's apparent confusion when the plexiglass was reversed emanated from her understanding of the concept. Because Nicky understood that placing the plexiglass over the objects would make the smaller object *appear* larger than the objectively larger object, she did not understand how when, after putting the plexiglass in place, the objectively larger object was even bigger, and the objectively smaller object remained the same size, and consequently she was unable to complete the task until the plexiglass was correctly oriented.

Intrasubject comparison on the Color and Size tasks. Figures 4.5 and 4.6 present the Color and Size Phase 4 intervention data and the Color and Size Task test data respectively in a manner that allows direct comparison between Nicky's performance in the two task domains. Figure 4.5 suggests that the Color tasks may have been nominally easier for Nicky to understand. Whereas Nicky's responses during Phase 4 of intervention on the Size tasks were consistently incorrect, there were some instances of comprehension of the Color teaching and generalization tasks, albeit inconsistent. In both task domains, seven intervention sessions were required before consistent, correct responding was established.

In looking at Figure 4.6, it is evident that consistent responding on tests of the Size and Color tasks began within a day of each other (Sessions 19 and 20 respectively). Thus, any difference in difficulty between the two domains was minimal. Figure 4.6 also suggests that

Color Tasks



Size Tasks

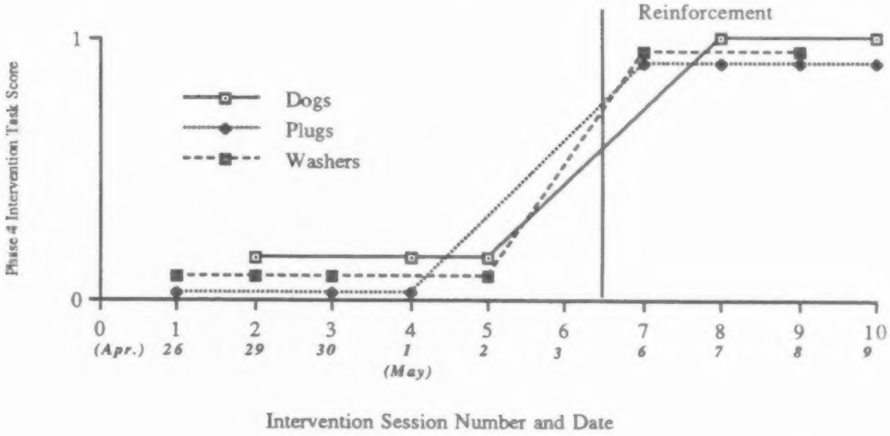
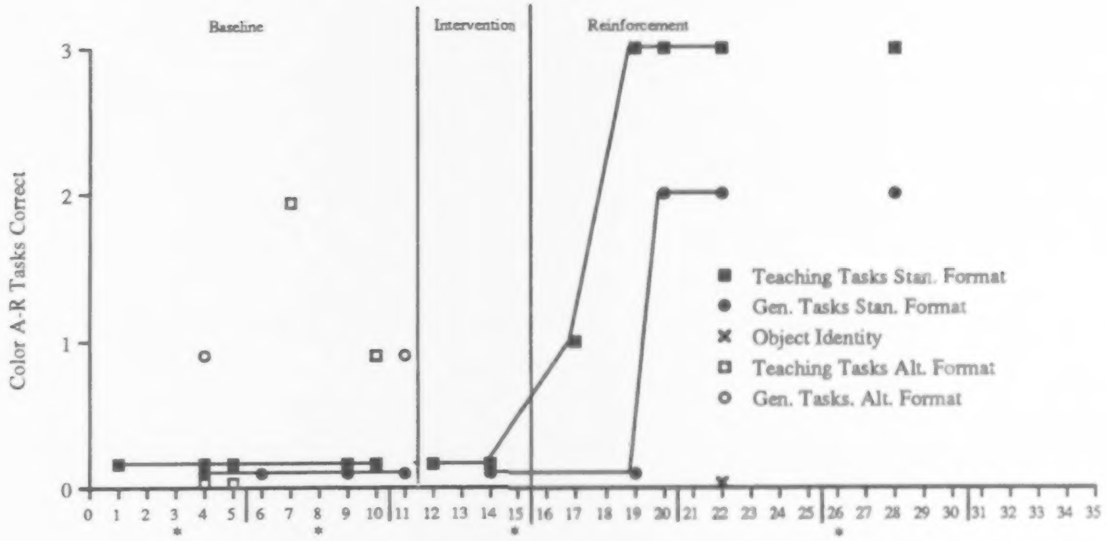


Figure 4.5. Nicky—Comparison of Phase 4 intervention on Color and Size tasks.

Color Tasks



Size Tasks

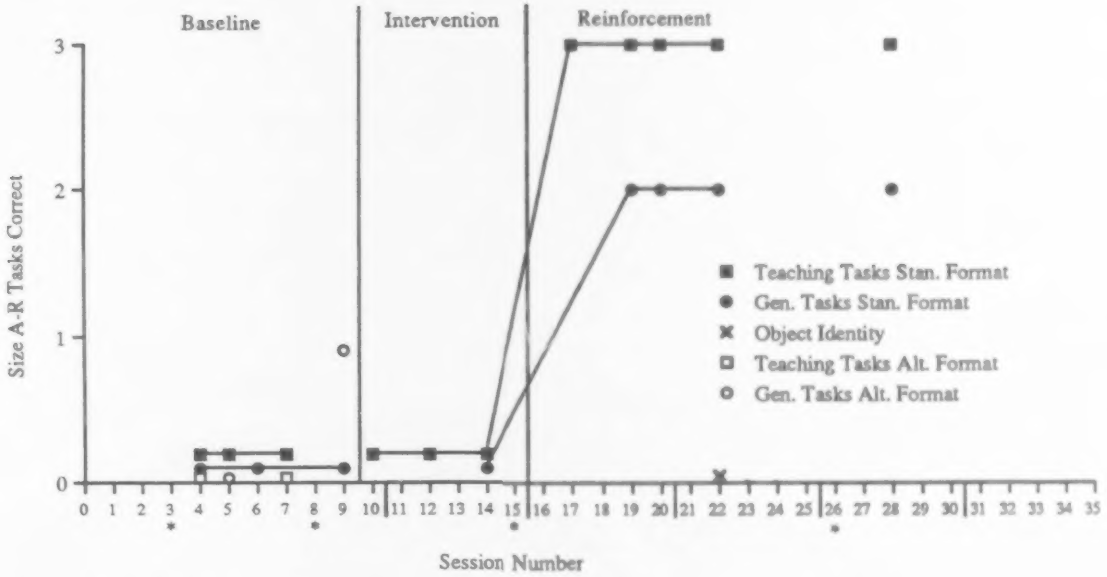


Figure 4.6. Nicky—Comparison of performance on tests of Color and Size tasks.

the introduction of secondary reinforcement had a positive effect on Nicky's ability to answer the Appearance and Reality questions correctly. After reinforcement was introduced in Session 16, Nicky immediately demonstrated mastery on the Size tasks, and, shortly thereafter (by Session 19), on the Color teaching tasks as well. The investigator believes that the use of secondary reinforcement motivated Nicky to pay closer attention to the instruction, and consequently achieve criterion on the tasks sooner than may have otherwise occurred. Because of the temporal proximity of Color and Size task mastery, the investigator concluded that either both of these task domains were of equal difficulty for Nicky to comprehend, or that instruction on one type of task was easily generalized to the other type of task.

Object Identity tasks. When presented with the first Object Identity task (Orange), Nicky was able to correctly identify the object as an orange when she saw it sitting on a mat on the opposite side of the table from her, in response to the first control question ("What is this?"). After Nicky was permitted to pick up the orange, she spontaneously said "plastic," and sniffed it before the investigator had a chance to ask the second control question ("What is it made of?") obviating the need to ask the question. After the orange had been replaced on the mat, the investigator asked the Appearance question ("When you first saw this over here like this, what did you think it was?"), to which Nicky correctly replied, "orange." However, when the investigator asked the Reality question ("What is it *really*?"), Nicky replied, "a lemon."

When the plastic egg was introduced to Nicky, she spontaneously identified it as an egg. Thus, the first control question was not asked. Prior to picking up the egg, Nicky announced that the egg was plastic. However, when the second control question (i.e.,"What is it made of?") was actually asked, Nicky did not respond. The investigator then said, "Tell me again. What is this?" Nicky then replied, "plastic." On this task Nicky was able to correctly answer both the Appearance and the Reality questions. However, when a repetition of the task was attempted using artificial flowers, Nicky, after correctly saying "flowers" in response to the first control question (i.e., "What is this?"), said that they were made of "plants" in response to the second control question. Nicky was apparently unaware that the flowers were artificial. This unawareness may be partially due to a lack of experience with flowers. Most nine year old children have had limited experience with flowers (beyond seeing them), artificial or otherwise.

After Nicky had reached criterion on the Color and Size tasks (Session 22), the Rock Object Identity task was administered. This task was rendered invalid, however, because Nicky touched it before she had answered the first control question, despite the direction to look first, and not to touch. The first control question was asked anyway, and Nicky replied that it was a rock. Nicky was then instructed to pick it up and handle it and was then asked the second control question, to which there was no response. The question was repeated, and again, Nicky did not respond. Other attempts were made to elicit a response from Nicky regarding the true nature of the sponge rock (e.g., "If you threw *this* rock at a window would *this* rock break the window?"), but the attempts were unsuccessful. Nicky did concede, however, that the "rock" was a pretend rock when asked, "Is this rock a pretend rock or a real rock?"

When the Orange Object Identity task was re-administered as a post-test, Nicky again did not seem to understand the requirements of the task. Although identifying the plastic orange as an "orange," in response to the question, "What is it made of?" Nicky said, "eating." Nicky was then asked if the orange was a pretend orange or a real orange and she said "pretend." The investigator then asked, "How do you know?" to which Nicky said, "After supper then I'll have to peel the orange. You have to make a glaze out of the orange juice." The investigator asked, "Could you do that with *this* orange (touching the orange)?," and Nicky replied, "Yes."

One can conclude from Nicky's responses on these tasks, that she either did not understand the requirements of the task, or was convinced that the imitation objects were, in reality, what they represented. Given Nicky's often irrelevant responses, however, the former explanation is the most likely one. Definitive conclusions are inadvisable since Nicky was unable to correctly answer the control questions.

Summary

To summarize, in terms of inter-rater reliability, Hartmann (1977) indicated that percentage agreement of 80% for trial reliability, and phi and kappa coefficients that exceed .60 are adequate for determining that an acceptable value of trial reliability has been achieved. Since the percent agreements and the phi and kappa coefficients of Nicky's data on both the Color and Size tasks exceeded these minimum standards, it may be concluded that trial reliability was achieved.

During baseline data gathering, Nicky consistently made phenomenist errors when answering the Reality questions on both the Color and Size tasks. Once intervention was introduced, criterion was reached on the Color teaching tasks by the 8th instructional session, on the Color generalization tasks by the 9th instructional session, on the Size teaching tasks by the 8th instructional session, and on the Size generalization tasks by the 10th instructional session. Nicky was unable to generalize her understanding of the appearance–reality distinction concept to the Object Identity task. It appeared that the introduction of the reinforcement procedure may have elicited better attention to the task at hand, since Nicky's performance on the Color and Size tasks reached criterion shortly after its introduction.

Although Nicky displayed uncertainty in her answers to the Reality questions during the initial stages of instruction in both task domains, once she appeared to understand the concept, her answers were consistent and correct, regardless of which question (i.e., the Appearance or the Reality question) was asked first. As will be seen, this uncertainty of response was seen in most of the other subjects as well.

David

Description

David is a member of a visible minority, whose family speaks both their native language and English in the home. Although born in a different country, David has been in Canada at least since the age of five, the age at which he was admitted into a local hospital's Behavior Preschool Program which he attended for two months. In the autumn of 1984, at the age of 5.0, David began attending the early education program at a local public school. In 1985, the public school board conducted an intellectual assessment of David. With a chronological age of 6.2, David achieved an overall intellectual age-equivalent of 3.9 on the Merrill Palmer test. Scores on the subtests, however, ranged from 18 months of age to 71 months. Thus, although David scored in the educable mentally handicapped range overall, his skills were quite scattered. The most severe area of weakness apparent during this assessment was in the area of language comprehension and usage. The school board also identified David as having autism (Autism Behavior Checklist) at the time of this assessment.

David is currently completely integrated into a regular grade 3/4 classroom where he has a modified program and the assistance of a teacher's aide.

Upon testing on the inclusion criteria for the present study, David (CA = 11.7) achieved a mental age-equivalent of 6.10 on the LIPS, a receptive vocabulary age-equivalent of 4.7 on the PPVT-R, and an MLU of 4.37. Hence, David was the highest scoring subject on the tests and had the longest MLU of all the subjects.

David was very cooperative throughout the study, and appeared to enjoy working with the investigator. He attended to the tasks being presented quite well, and only occasional prompts to listen or look at the presentation were required. The most noticeable behavior exhibited by David was his tendency to mumble. Although the investigator could not always hear what he was saying, those mumblings that could be distinguished seemed to be remarks unrelated to the situation. On other occasions David made very clear irrelevant remarks. For example, when David was asked the Appearance question on the Egg Object Identity task ("When you saw this over here, what did you think it was?"), David replied, "It's a elevator." However, when the investigator prompted him not to be silly and repeated the question, David said, "It's a egg." Despite the linguistic complexity of the Appearance question in the Object Identity tasks, it seems clear that this did not contribute to David's making his irrelevant comment, since he answered correctly when prompted.

Inter-Rater Reliability

Color tasks. Table 4.5 presents the inter-rater reliability data for David on the Color tasks. On 50 observations, there were no disagreements regarding behavior occurrence or nonoccurrence. There were 31 agreements between the observers of behavior occurrence, and 19 agreements of behavior nonoccurrence. Thus, total percent agreement, effective percent agreement on behavior occurrence, and effective percent agreement on behavior nonoccurrence were all 100. The resulting phi and kappa coefficients were necessarily 1.0.

Table 4.5

David—Inter-rater Reliability on Color Tasks

		Observer 2		
		0	1	
Observer 1	1	A 0	B 31	A + B = 31
	0	C 19	D 0	C + D = 19
		A + C = 19	B + D = 31	N = 50

Size tasks. As seen in Table 4.6, of 45 observations, there were 27 agreements of behavior occurrence between the observers, and 18 agreements of behavior nonoccurrence. There were no disagreements between the observers. Thus, as in the Color tasks, total percent agreement, effective percent agreement on behavior occurrence, and effective percent agreement on behavior nonoccurrence were also 100 for the Size tasks, and the resulting phi and kappa coefficients were 1.0.

Table 4.6

David—Inter-rater Reliability on Size Tasks

		Observer 2		
		0	1	
Observer 1	1	A 0	B 27	A + B = 27
	0	C 18	D 0	C + D = 18
		A + C = 18	B + D = 27	N = 45

Results

Color tasks. As seen in Figure 4.7, when the Appearance and Reality questions were asked using the standard format, David was unable to complete the task correctly during baseline measures, with one exception, that may be attributed to chance. However, when the alternate question format was used, David had more success on the Color A-R tasks. As Table 4.7 illustrates, of the 7 trials in which the alternate format was used, David was able to answer the questions correctly on 6 of them (86% success). Thus, for David, the question format used made a distinct difference in his ability to complete the tasks correctly.

Perhaps the clearest illustration of the linguistic difficulties inherent in A-R distinction tasks as specified in Chapter Two, is seen in David's answers to the Reality questions when the standard question format was used during baseline. On 12 of 16 trials in which the standard question format was used, David, on being asked, "What color is it (the yellow koala) really?," replied that it was "*really* (red)." Although this answer in itself does not signify a misunderstanding of the question's intent, David's voice inflections seemed to indicate that he was mistaking the Reality question ("What color is it really?") as an enquiry as to the intensity of the object's color, rather than an enquiry as to the object's veridical color.

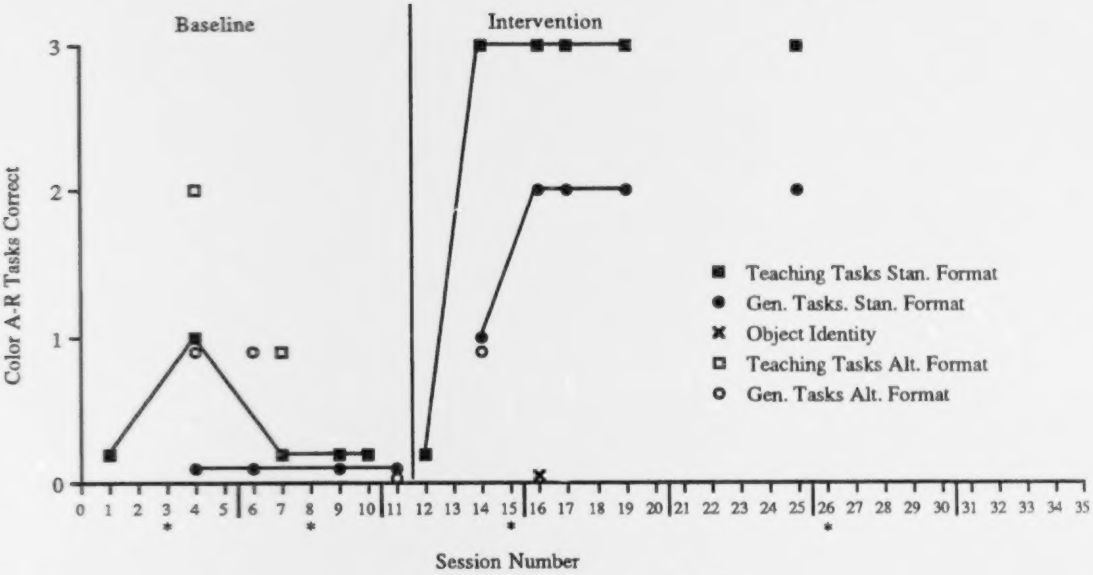


Figure 4.7. David—Number of correct responses on tests of Color tasks across conditions.

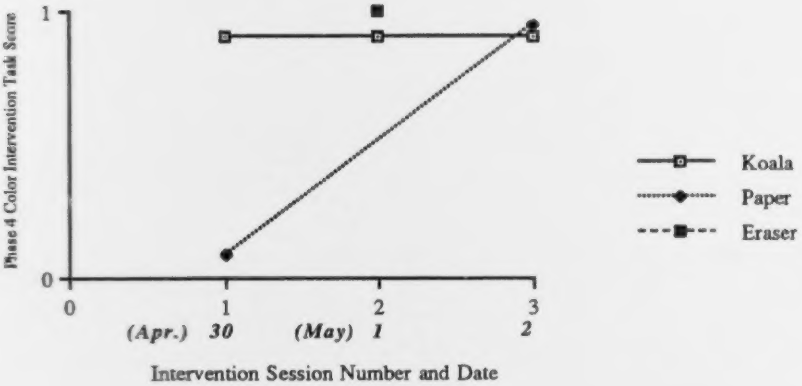


Figure 4.8. David—Phase 4 intervention on Color tasks.

Table 4.7

David—Data Breakdown of Color Sessions When Alternate Question Format Used

		Standard Format		Alternate Format	
Session Number	Task	Pass	Fail	Pass	Fail
4	Paper	√			
	Eraser			√	
	Koala			√	
	Crayon (CG)		√		
	Chalk (CG)		√		
	Chalk (CG)			√	
6	Crayon (CG)			√	
	Chalk (CG)		√		
7	Paper		√		
	Eraser		√		
	Koala			√	
11	Chalk (CG)				√
	Crayon (CG)		√		
14	Eraser	√			
	Paper	√			
	Chalk (CG)			√	
	Crayon (CG)	√			
	Koala	√			

This kind of answer was also evident during instruction, as seen in the following videotape excerpt of the latter half of Phase 2 of the Koala task in the first instructional session.

I: (Put yellow koala into container such that the entire object appeared red) *Really* it's yellow. What color is it *really?*

D: Red.

I: (Takes koala out of container) *Really* it's yellow. What color is it *really?*

D: Yellow.

I: (Replaces Koala in container) What color is it really?

D: Yellow.

I: But it *looks* red. What color does it look?

D: Really red.

I: It looks red, but really and truly it's yellow.

David's rapid acquisition of the A-R distinction after instruction was initiated, supports the interpretation that language, rather than understanding of the A-R distinction concept, was responsible for David's phenomenist mistakes. As seen in Figure 4.7, David demonstrated his understanding of the concept (as applied to the property of color), in both the teaching tasks and the generalization tasks, when he was tested in the third intervention session (Session 14). The rapidity of David's acquisition of understanding is even more clearly seen in Figure 4.8. This figure demonstrates that by the second intervention session, David was able to correctly complete Phase 4 of the instructional procedure for both tasks presented. Thus, if David had a theory of mind deficit, and truly did not understand how things could look one way but really be something else, it is unlikely that he would have acquired that understanding after only two periods of instruction. The more likely explanation is that the instructional periods, in which the investigator provided a model of task completion, allowed David to determine the task requirements, and subsequently complete the tasks correctly and independently.

Before understanding of the the A-R tasks was exhibited during instruction, David's behavior was very similar to that of Nicky. In the first instructional session, David apparently believed that the color an object looked at any given time, was the object's true color. For example, David agreed that an object was "really" one color (the object's true color), each time he was asked while the filter gradually covered the object. Once the object was completely covered by the filter, however, David changed his answer (even when being asked the same question), and stated that the object was now "really" the color of the filter.

In fact, before instruction was initiated, David's comments seemed to indicate that he believed that the filter actually changed the color of the object. For example, on the Paper task in the first baseline session, David correctly answered the control question ("What color is this?" [investigator pointed to the paper before the filter was placed over it]). However, as soon as the filter was placed over the paper, David spontaneously said, "Changed into a green." He then went on to complete the task, giving a phenomenist answer to the Reality question. On the Chalk task in Session 4 David removed the filter himself and announced that "It [the chalk] turned back into a white." Thus, although aware that the chalk was white underneath the filter, David's choice of words seemed to indicate that he believed that the filter *actually* changed the color of the object. This behavior, however, does not negate the possibility that language was the primary difficulty inhibiting David's understanding of the appearance-reality distinction tasks during baseline and the first instructional session. It is possible that David's difficulty with language usage resulted in inappropriate word use (e.g., "It turned back into a white.") rather than non-understanding of the concept itself. If this explanation is true, the intervention sessions probably served to clarify the terminology for David, rather than instruct him in the A-R distinction concept.

Other similarities to Nicky's behavior included David's interest in using the filters to alter the color of his hands and clothing during the intervention sessions, and his tendency to change his answer as soon as the investigator began to take away the filter. For example, during Phase 4 of the Paper task in the first intervention session, David correctly covered the white paper with the green filter when asked to use the filter to make the paper *look* green. When asked, "What did you do?," David said, "Made it green." The investigator corrected David saying, "You made it *look* green. But what color is it *really?*," to which David replied, "green." When the investigator started to peel away the filter, David changed his answer to white. The investigator then modelled the task again to David and repeated Phase 4. This time, after stating that the real color of the paper was green, David immediately changed his answer to white as soon as the investigator made a move to remove the filter. It is probable that the investigator's movement to remove the filter at this stage of instruction served as a gestural prompt for both Nicky and David to alter their answers. However, once the requirements of the task were understood, the external prompt was no longer necessary—Nicky and David were able to complete the Color A-R tasks correctly without the provision of the prompt.

To conclude, because David demonstrated understanding of the Color A-R tasks so quickly after instruction was introduced, the investigator believes that difficulties with language comprehension and usage, rather than a theory of mind deficit, accounted for David's inability to complete the tasks correctly when they were first presented.

Size tasks. When David was first presented with the Size tasks, it was apparent that he either did not have a firm understanding of the concept "bigger," or that he enjoyed having the investigator believe that he did not have a firm understanding of the concept. For example, although David was able to correctly identify which dog was larger at each stage of the Dog task in Session 4, and in fact, passed the task, the remainder of the Size tasks presented during the session were unscorable, since David did not respond in a consistent correct manner to the instruction to "point to the bigger one," even when the plexiglass was not over the objects (thus distorting the true size of one of them). The investigator then used the materials of the "bigger" tasks (file cards, notepads, batteries, paperclips, and stickers of dogs) to determine if David would respond consistently to these materials when asked to point to the larger object. However, a consistent response was not established. When the investigator asked David to point to the smaller object in various presentations of the "bigger" task materials (to determine if he would respond reliably to that instruction), David's responses remained inconsistent.

Consequently, during the next session in which the investigator worked with David (Session 6), the investigator attempted to teach the concept "bigger" using Direct Instruction (DI) techniques and the materials of the "bigger" test. Despite this instruction David's responses were inconsistent. After 6 DI trials using different materials, the investigator prompted David to look at each object carefully and to "point to the bigger one," to which David responded by putting his head close to the objects on the table and pointing to the small object. After three trials of this behavior, the investigator prompted David firmly saying, "Don't be silly. Just point to the one that's bigger," whereupon David lifted up his head, and correctly pointed to the larger object on this, and all subsequent trials of the "bigger" tasks for the remainder of the study. It is likely, then, that David understood the concept of bigger before instruction began, but enjoyed the additional attention that the instruction provided. To ensure that David's identification of a "bigger" object remained consistent, concept understanding was tested again at the beginning of Sessions 7 and 9. On a total of 20 trials in these sessions, David made one mistake on the 9th trial. Although he made phenomenist errors in the course of the A-R task administration, David was able to consistently identify the larger object before and after the plexiglass covered the objects.The investigator concluded that David's responding was consistent enough to proceed with the study.

As can be seen in Figure 4.9, once consistent responding was established, a stable baseline was easily obtained. David was unable to complete any of the Size tasks correctly, and consistently made phenomenist errors.

Table 4.8 and Figure 4.9 illustrate that the use of the alternate question format did not assist David in completing the Size A-R tasks correctly during baseline. On the two occasions in which the alternate format was employed, David was unable to complete the Size A-R tasks correctly.

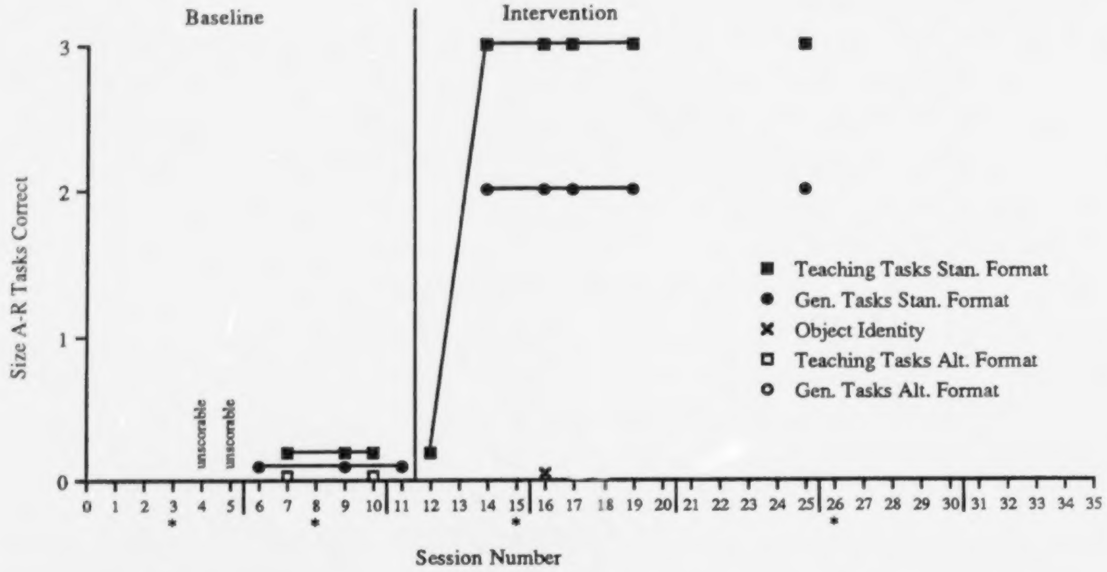


Figure 4.9. David—Number of correct responses on tests of Size tasks across conditions.

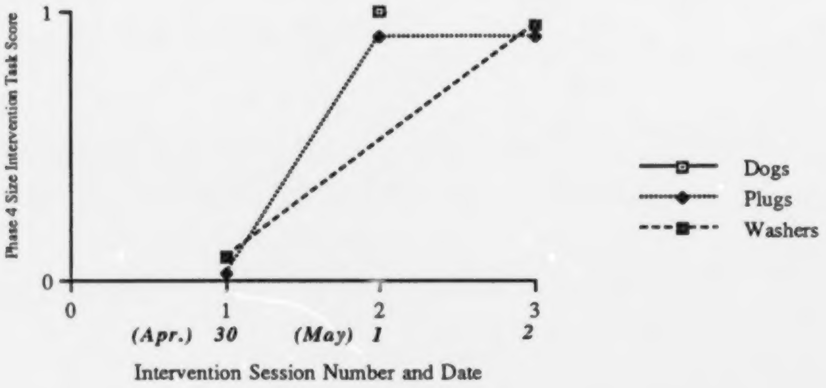


Figure 4.10. David—Phase 4 intervention on Size tasks.

Table 4.8

David—Data Breakdown of Size Sessions When Alternate Question Format Used

		Standard Format		Alternate Format	
Session Number	Task	Pass	Fail	Pass	Fail
7	Washers		√		
	Dogs				√
	Plugs		√		
10	Washers		√		
	Plugs		√		
	Dogs				√

However, once intervention was introduced, David displayed his understanding of the tasks by the second testing session (Session 14 in Figure 4.9) completing all tasks correctly on this and all subsequent testing sessions, including the maintenance testing session. Similar to David's performance on the Color tasks, only two intervention sessions were required before David was able to complete Phase 4 correctly on the tasks presented (Figure 4.10).

Although David imitated the investigator's use of the acetate correctly when it was used in the first intervention session, its use encouraged David to make remarks unrelated to the situation. For example, in Phase 2 of this session on the Washers task, David commented, "That's the way we go on the subway," while he traced the lines. Although the lines may have resembled subway tracks to David, it was obvious that they did not assist David in "proving" which object was bigger. The use of the acetate was discontinued after this session.

Despite David's ultimate failure on Phase 4 of the Washer task in the first session, it was evident that he had some understanding of the A-R concept. Specifically, after the investigator had modelled the correct responses to the Appearance and Reality questions using the acetate, and prior to David's comment about the subway, David said, "But really bigger is the small one." This paraphrasing of the investigator's model ("But *really* this one is bigger"[pointing to the objectively larger object after the plexiglass was removed and the acetate was put in place]), seemed to indicate that David realized that although one object appeared bigger when the plexiglass was placed over the objects, it was the other object that was really bigger.

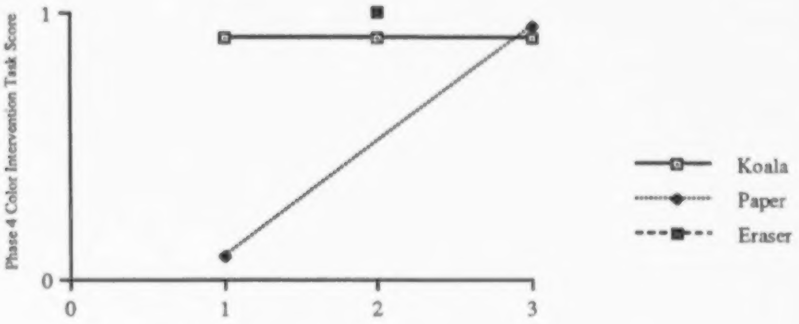
David demonstrated his understanding of the Size A-R tasks very quickly in the second intervention session. The Dog task was the first one presented in this session. All four phases of this task were presented and completed correctly, with only one presentation of each phase being necessary. After Phase 4 had been completed, David took the plexiglass and placed it over the dogs such that the smaller dog was magnified and appeared larger than the objectively larger one. Pointing to the magnified dog, David commented, "On the other hand it looks bigger and on the other hand (pointed to the phenomenally smaller dog) it's really bigger. (Removed plexiglass) There are two puppy dogs. The same (points to bigger one) and a bigger one (points to smaller one)." The investigator asked, "Which one's the bigger one?," and David correctly pointed to the larger dog. Although David's pointing after his removal of the plexiglass did not correspond with his verbalizations, the investigator believed that David did understand the concept, but simply made a mistake in his pointing. This belief was confirmed by the investigator's verification of David's knowledge of which dog was bigger when the plexiglass was not placed over the dogs.

David's apparent misunderstanding of the term "really" that was witnessed in the Color tasks, was not evident during instruction or testing in the Size tasks. This may be because the Size tasks did not necessitate a verbal response. The children were simply required to point to the object that either "looked" bigger, or was "really" bigger in a given task. Thus, although David may have interpreted the Reality question of the Size tasks during baseline as an enquiry as to the degree of size difference between the objects, his verbalizations did not make this interpretation evident to the investigator.

Intrasubject comparison on the Color and Size tasks. In comparing David's performance during the Size and Color intervention sessions (Figure 4.11), it is evident that these different task domains were generally equivalent in difficulty for David to comprehend. Consistent correct responding during Phase 4 of instruction was established in both task domains by the second session. Similarly, as seen in Figure 4.12, once consistent responding was established in the instructional sessions, consistent correct responding was also established when both task domains were tested in Session 14, and in each subsequent testing session. Understanding of both task domains was also maintained after eight days of non-intervention.

Concrete reinforcement was considered unnecessary for David since he attended to the tasks well, and achieved criterion within two sessions. Hence, only verbal praise and encouragement was used to maintain David's motivation and interest in the tasks.

Color Tasks



Size Tasks

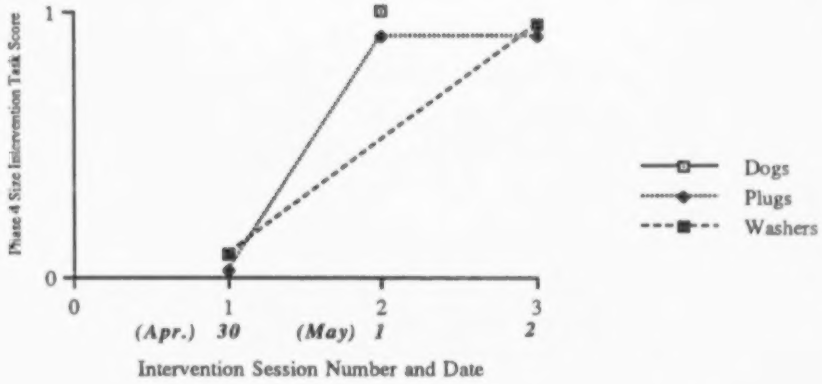
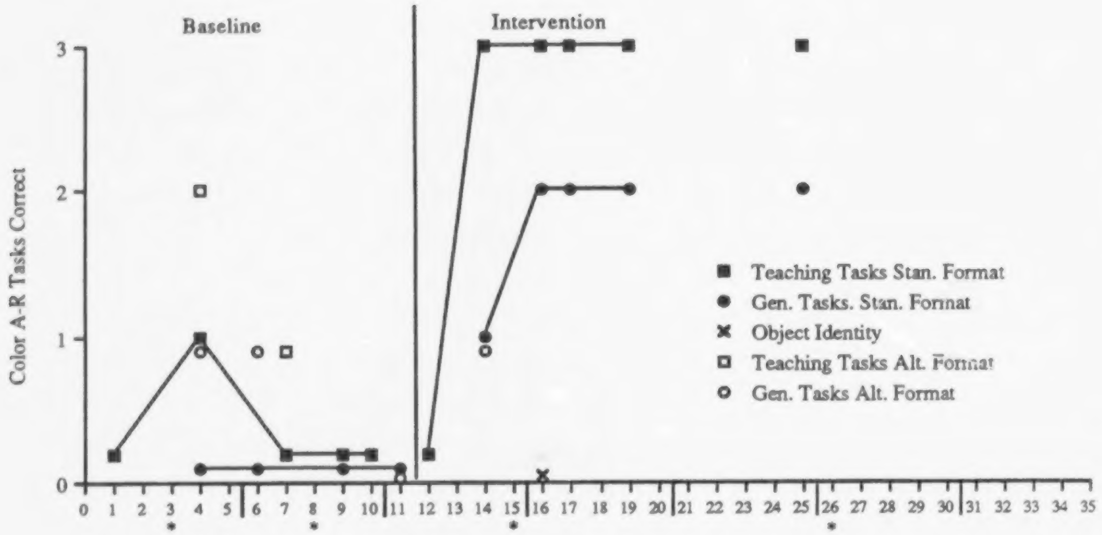


Figure 4.11. David—Comparison of Phase 4 intervention on Color and Size tasks.

Color Tasks



Size Tasks

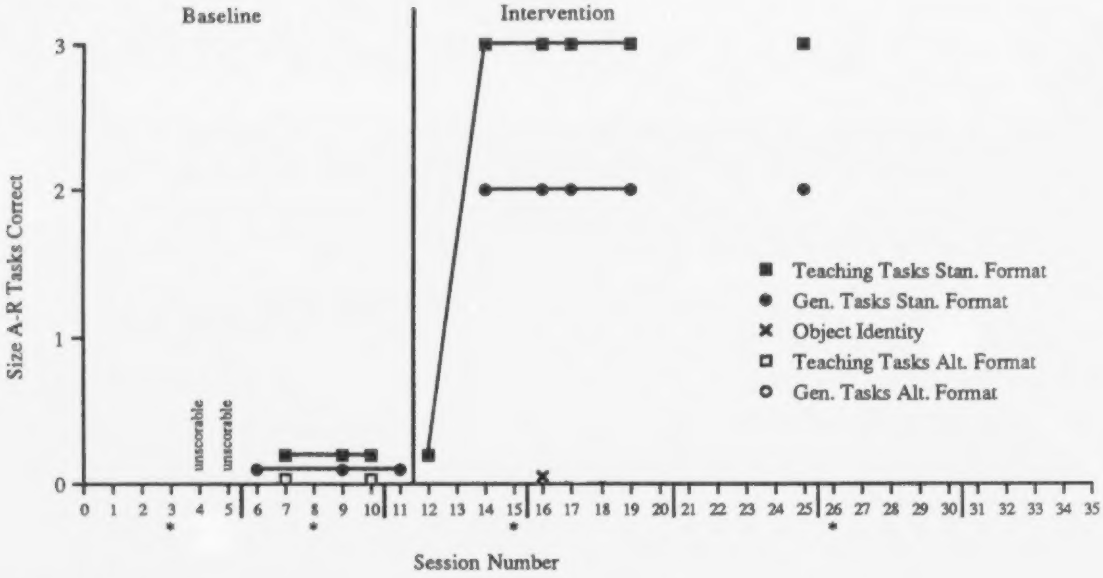


Figure 4.12. David—Comparison of performance on tests of Color and Size tasks.

Object Identity tasks. Although David's answers on the Object Identity pretest (the plastic orange) provided some indication that he understood the difference between the orange's appearance and its reality, the investigator inadvertently failed to ask the second control question ("What is it made of?"). Thus, a firm conclusion regarding David's understanding of the Appearance-Reality distinction as it relates to the Orange Object Identity task cannot be reached. However, David's performance on other object identity tasks suggested that this task domain was not understood. After the investigator instructed David to pick up the orange, David said, "It's a ball," and rolled it on the table. Because David's remark was mumbled, the investigator enquired, "What is it?," to which David replied, "Orange." The investigator then asked, "Is it a real orange?" and David replied, "A ball." The investigator then proceeded with the task, replacing the orange to its original location and asking the Appearance and Reality questions. In response to the Appearance question ("When you first saw this over here like this, what did you think it was?") David said, "It was orange," and in response to the Reality question ("What is it really?") David said, "It's a ball."

Although David's remarks and responses during the administration of the Egg Object Identity task were usually *related* to the object that was being used (that is, his comments referred to things that could be done with an egg), on the whole, they did not seem to be directed at the question that had been asked. This problem is demonstrated in the following videotape excerpt.

I: (Reveals plastic egg). What is this?
D: This is egg. Scramble up the eggs.
I: Touch it.
D: (Touches egg)
I: What is it made of?
D: It's butter egg pizza.
I: What is it made of David?
D: Yellow.
I: Is it a real egg or a pretend egg?
D: Pretend egg.
I: (Replaces egg on far side of table from subject) When you first saw this over here like this, what did you think it was? (Appearance question)
D: It was a boiled.
I: Listen. When you first saw this over here like this, what did you think it was?
D: Cooking egg and ham.
I: Listen to the question. Pay attention.

D: Think it was?

I: Yes. What did you *think* this was?

D: It's a butter.

I: What did you think this was when you saw this over here?

D: It's a elevator.

I: Without being silly David. What did you think it was?

D: It's a egg.

I: And what is it really?

D: It's really dip it in a bowl.

The investigator believes that David's responses to this task, and to the Object Identity task using the artificial flowers, indicated his incomprehension of the task requirements. On the Color and Size tasks, tasks that appeared to be easier for all the subjects to understand and complete, David attended better, and refrained from making irrelevant remarks.

As illustrated by the following videotape excerpt, the comments that David made when the Rock Object Identity task was administered after criterion had been reached on the Color and Size tasks, confirmed the investigator's belief that the Object Identity task requirements were not completely understood by David.

I: (Reveals "rock" placed on mat) What is this?

D: Rock

I: Pick it up.

D: It's a hand.

I: What is it made of?

D: Rock

I: Is this a pretend rock or a real rock?

D: Pretend

I: Would you hurt someone if you threw this rock at them? (indicating rock).

D: No.

I: Why not?

D: 'Cause I like them.

I: Let's say you accidently threw a rock but somebody got in the way. Would this rock hurt them?

D: No.

I: Why not?

D: 'Cause he loves this rock.

I: Why wouldn't *this* rock hurt them?

D: 'Cause this feels nice soft (rubbing the rock).

I: (Replaces rock to original location) When you first saw this over here like this, what did you think it was?

D: A rock.

I: What is it really?

D: Really a rock.

Similar responses were also seen on the Orange Object Identity post-test:

I: (Reveals "orange" placed on mat) What is this?

D: Orange

I: O.K. You can pick it up.

D: (Picks up orange, bounces it on table and tosses it from hand to hand) It's a ball.

I: (Replaces orange on mat) When you first saw this over here like this, what did you think it was?

D: It's an orange.

I: What is it really?

D: Really orange.

These excerpts show that although David was aware that the "rock" and "orange" being used in these tasks were not real, he was apparently unable to determine exactly what they *were* made of, at least when he was asked in the format used in this study. Although the communicative intent of the investigator's Appearance and Reality questions implied that different answers to each question were required, David persisted in saying that the rock and orange both *looked* like a rock and orange, and *were* these objects in reality as well. However, the reader must keep in mind the linguistic difficulty inherent in this task. This difficulty is discussed more thoroughly in the Discussion chapter.

Summary

In summary, the results of the inter-rater reliability calculations indicated that both observers agreed on the behavior occurrence and nonoccurrence of David's responses 100% of the time. Thus, trial reliability was obtained, and interpretations of the data are possible.

During baseline task administration, David consistently made phenomenist errors on the Color and Size appearance-reality tasks when the standard question format was used. However, when the alternate question format was used in the Color tasks, David was able to complete 6 out of 7 items. No similar difference in baseline performance was noted in the Size tasks when the alternate format was used.

David reached criterion on both the Color and Size appearance-reality tasks after only two instructional periods. Instruction on these tasks began on the same day (Session 12), and testing in Session 14 indicated that mastery of both task domains had been achieved.

There was some indication during instruction on the Color tasks, that David was interpreting the Reality question as an enquiry as to the intensity of the apparent color rather than as an enquiry as to the true color of the object. However, after the expected answer was modelled a few times by the investigator, David was able to complete the tasks correctly.

Similar to Nicky, David was unable to generalize his understanding of the appearance-reality distinction concept to the Object Identity tasks. Although his answers on these tasks indicated his awareness that the objects were not really what they represented, he remained unable to complete the tasks correctly.

Brian

Description

Brian was identified at the age of 3.4 by a local hospital, as having many characteristics of autism. A Merrill Palmer test administered at that time determined that Brian's IQ was low average. At the age of 3.9 the local school board determined that Brian met its criteria for autism, and admitted him into an early education program. He also received occupational and speech therapy at the hospital where diagnosis had been made. A LIPS and a Stanford-Binet Form L-M, administered when Brian was five years of age, found that he was functioning at the high end of the educable mentally handicapped category.

Brian is currently enrolled in a regular grade four classroom where he has an individualized program and the assistance of a full-time aide. Although he lives in a group home during the week, Brian goes home for the weekends.

Brian was visibly upset during his initial meeting with the investigator. He came into the room willingly, bringing his crayons and coloring book, but screamed, hit himself by tapping his forefingers against his chin and hitting his chest, drummed his fingers on the table, and was echolalic much of the time. For example, in a loud voice he said things like, "It is all right, it is all right!" "Jill [his assistant] is coming back! Jill is in the washroom!" "It is Sesame Street! It is the meaning! It is the meaning! It is Sesame Street!" "It is getting late!" "Calm down!" and "No need to shout!" During the initial 10 minutes, he kept his back to the investigator. After the investigator began drawing on a piece of paper, verbalizing what she was drawing, Brian immediately stopped shouting and hitting himself, but remained with his back turned. At first Brian just glanced at what the investigator was doing, but then watched carefully. After a couple of minutes of watching, Brian opened his own coloring book and began drawing many clouds (imitating the investigator). He drew quietly, and responded to the investigator's suggestion of counting all the clouds by counting them aloud. Brian continued to draw until recess (about 15 minutes), for which he was permitted to go outside. Brian returned to the room with the investigator after recess;

pretesting then proceeded without further disruption. Brian went willingly and calmly with the investigator for the remainder of the study, and there were no further instances of the behavior seen in the first session.

Brian's chronological age at the beginning of the study was 9.8. On the current measures of the PPVT-R and the LIPS, Brian received age-equivalent scores of 5.6 and 4.5 respectively. Thus, Brian was the only subject whose nonverbal mental age was lower than his receptive vocabulary age-equivalent.

Brian's MLU on the language sample was 1.89. Thus, Brian's MLU was the lowest of all the subjects. However, he was also the shyest subject, and had the most difficulty adapting to new people, as evidenced by his behavior the first time the investigator met with him. Although the investigator heard Brian using complete sentences when talking to his aide, it was extremely difficult to elicit more than single word utterances for the language sample, or indeed, at any point during the study. Thus, the current estimation of Brian's MLU is not considered to be a valid indication of his expressive language capability.

In the course of the study, Brian exhibited numerous instances of echolalic speech and behavioral perseverations. He tended to echo the investigator's questions when it appeared that he did not understand them. For example, although clearly understanding the concept of bigger as tested by the "bigger" test, when asked during the Size task, "Is one bigger?," or "Which one is bigger?," Brian simply repeated the question without pointing to either object. However, when asked to "Point to the one that's bigger." Brian correctly pointed to the objectively larger object (when the plexiglass was not used). This response seemed to indicate that it was the question format with which Brian was having difficulty, not the implicit concept of the questions. As discussed in the previous chapter, the statement "Point to the one that's bigger" was substituted for the question "Which one is bigger," since the former appeared to be more easily understood.

Periodic perseverative responses were identified during testing on the Size tasks. On occasion, Brian would point to the left object in response to the direction to point to the larger object, regardless of whether the larger object was on the right or left, or whether the plexiglass sheet was over the objects or not. This response was considered perseverative only when he was asked to point to larger object *without* the plexiglass in place (i.e., the objects' true sizes were evident without being distorted in any manner) and a constant position bias was exhibited. When a perseverative response was suspected, the response was considered unscorable. Confirmation of perseveration came when the "bigger" tasks were administered and perseverative responses were detected in this exercise as well. When perseveration occurred during the "bigger" tasks, the investigator held up her index finger in front of Brian (gestural prompt), firmly prompted him to listen, and repeated the question after the objects had been rearranged. This verbal/gestural prompt usually resulted in a nonperseverative response. If correct responses were obtained on a number of trials, the Size task was re-administered. If, however, perseveration continued, testing was discontinued.

Inter-Rater Reliability

Color tasks. As seen in Table 4.9, out of a total of 47 observations, there was one instance of disagreement in which the investigator recorded an occurrence of behavior, while the second observer recorded a nonoccurrence (Cell A). There were 21 agreements of occurrence, and 25 agreements of nonoccurrence.

Table 4.9

Brian—Inter-Rater Reliability Color Tasks

		Observer 2		
		0	1	
Observer 1	1	A 1	B 21	A + B = 22
	0	C 25	D 0	C + D = 25
		A + C = 26	B + D = 21	N = 47

Total percent agreement on the Color tasks was 97.87, effective percent agreement on occurrence was 95.45, while effective percent agreement on nonoccurrence was 96.15. Coefficient calculations resulted in a phi coefficient of .96 and a kappa coefficient of .96 as well.

Size tasks. Inter-rater reliability scores of the Size tasks for Brian are presented in Table 4.10. There was one disagreement where the investigator recorded an occurrence of behavior, and the second observer recorded a nonoccurrence (Cell A), five occurrence agreements, and 50 agreements on behavior nonoccurrence on a total of 56 observations.

Total percent agreement on the Size tasks was 98.21, effective percent agreement on occurrence was 83.33, effective percent agreement on nonoccurrence was 98.04. The phi coefficient was .90 and the kappa coefficient was .89.

Table 4.10

Brian—Inter-Rater Reliability on Size Tasks

		Observer 2		
		0	1	
Observer 1	1	A 1	B 5	A + B = 6
	0	C 50	D 0	C + D = 50
		A + C = 51	B + D = 5	N = 56

Results

Color tasks. Brian's performance on the Color appearance-reality distinction task is presented in Figure 4.13. A stable baseline was easily obtained, with Brian consistently stating that the object's apparent color after being altered by a colored filter, was the object's true color. Thus, Brian made only phenomenist errors on the Color A-R tasks during baseline. The order in which the Appearance and Reality questions were asked did not affect the type of response that Brian gave. That is, regardless of which question was asked first after the filter was in place (the Appearance question or the Reality question), Brian replied that the apparent color of the object was, in fact, the true color of the object.

Use of the alternate question format also had no effect on Brian's ability to complete the Color A-R tasks correctly. As seen in Table 4.11 (and Figure 4.13), of the three items that were administered using the alternate format, Brian was unable to complete any correctly.

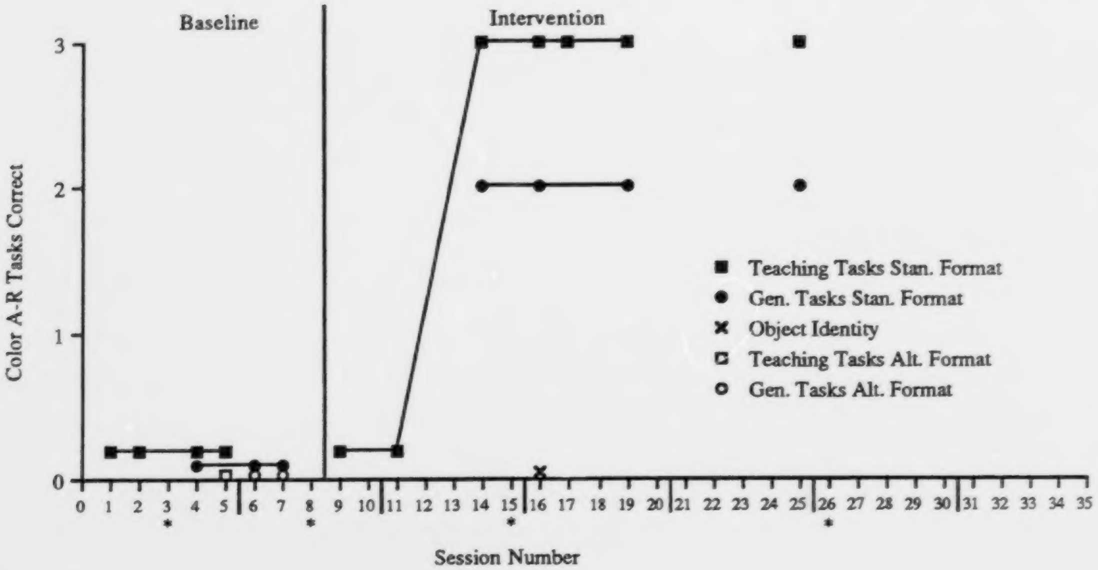


Figure 4.13. Brian—Number of correct responses on tests of Color tasks across conditions.

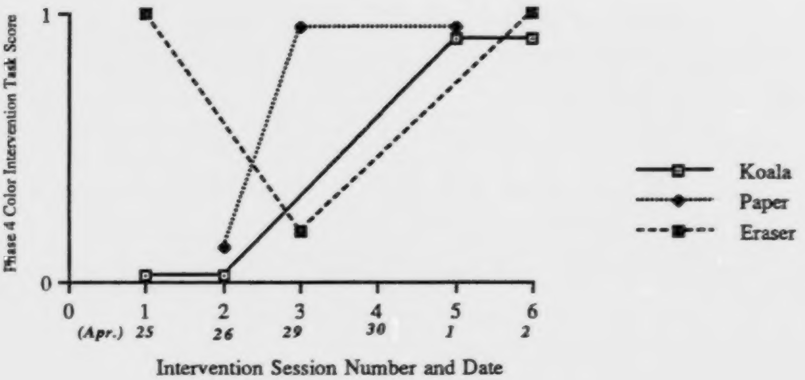


Figure 4.14. Brian—Phase 4 intervention on Color tasks.

Table 4.11.

Brian—Data Breakdown of Color Sessions When Alternate Question format Used

		Standard Format		Alternate Format	
Session Number	Task	Pass	Fail	Pass	Fail
5	Koala		√		
	Paper				√
	Eraser		√		
6	Crayon (CG)				√
	Chalk (CG)		√		
7	Crayon (CG)				√
	Chalk (CG)		√		
	Crayon (CG)		√		

Acquisition of the Color task, however, was fairly rapid after the intervention sessions were initiated without the use of secondary reinforcement. Examination of Figure 4.13 reveals that Brian was able to complete both the instructional Color A-R tasks and the color generalization tasks correctly when tested in Session 14. Consistent and correct responding on all tasks was maintained for the remainder of the study, including the maintenance check, administered eight days after intervention had been completed.

Data from Phase 4 of the instructional sessions is presented in Figure 4.14, and shows that a consistent and correct response to Phase 4 of the intervention was established by the fifth intervention session. Although responding was somewhat inconsistent during the initial sessions, the investigator believes that Brian did possess some concept of the appearance-reality distinction, but he may not have demonstrated it because of the investigator's method of task presentation during the first instructional sessions. For example, whenever the investigator modelled the correct answer to the Reality question (e.g., "It's *really* white"[in the case of the Paper task]), the filter was removed before the "real" color of the object was mentioned. When Brian was required to complete Phase 4, however, he was asked the Reality question while the filter was still in place. Thus, Brian was not provided with an exact model of what was required until the third intervention session. As seen in Figure 4.14, Brian was able to complete Phase 4 correctly in the next intervention session (because

Brian's language sample was obtained on April 30, no instruction took place in Session 4). It may be, then, that had an appropriate model been presented in an earlier intervention session, Brian would have been able to complete Phase 4 correctly sooner.

When testing Brian in Sessions 16 and 19, after understanding of the Color A-R concept had been demonstrated, the investigator found that when the Reality question was asked first, Brian responded with the apparent color of the object concerned. However, when the same task was re-administered in the same session, with the Appearance question being asked first, Brian completed the task correctly. It was apparent that Brian had become accustomed to answering the Appearance question first, and did not really listen to the question being asked. In fact, the Reality question was not asked first on any task between Sessions 9 and 16. However, after prompts to listen carefully were given in Session 19, and the tasks were repeated, Brian was able to complete all tasks correctly regardless of the question order, both in this session, and in the maintenance session eight days later. It is interesting to note that on all the testing occasions subsequent to mastery on the Color tasks, and during Phase 4 on tasks that were completed correctly, Brian would, after correctly answering the Reality question, spontaneously remove the filter, seemingly to verify his answer.

In summary, Brian had little difficulty in reaching criterion on the Color A-R tasks. Because of Brian's speed in demonstrating understanding of the concept (only 5 intervention sessions), the investigator concluded that language comprehension, rather than a deficit in understanding the appearance-reality distinction probably accounted for Brian's initial misunderstanding of the task. As noted above, once a true model was provided (i.e., the "true" color of the object concerned was modelled by the investigator *while* only its apparent color was visible), Brian quickly demonstrated understanding of the task, and even verified his answer by removing the filter (or pulling the object out from behind the filter).

Size tasks. As can be seen in Figure 4.15, there was somewhat more difficulty in obtaining a baseline for Brian on the Size tasks compared to baseline gathering on the Color tasks. This was because of Brian's tendency to perseverate (e. g., Session 4) and alternate his answers (e.g., Session 9) regardless of the question being asked. On each scorable session, Brian made both realist errors (indicated through pointing, that the objectively larger object was also the phenomenally larger object when the plexiglass was in place) and phenomenist errors (indicated that the phenomenally larger object was also the objectively larger object when the plexiglass was in place).

Because Brian did not answer the questions on the Size pretest task consistently, the "bigger" tasks were first administered in Session 2 to confirm his understanding of the

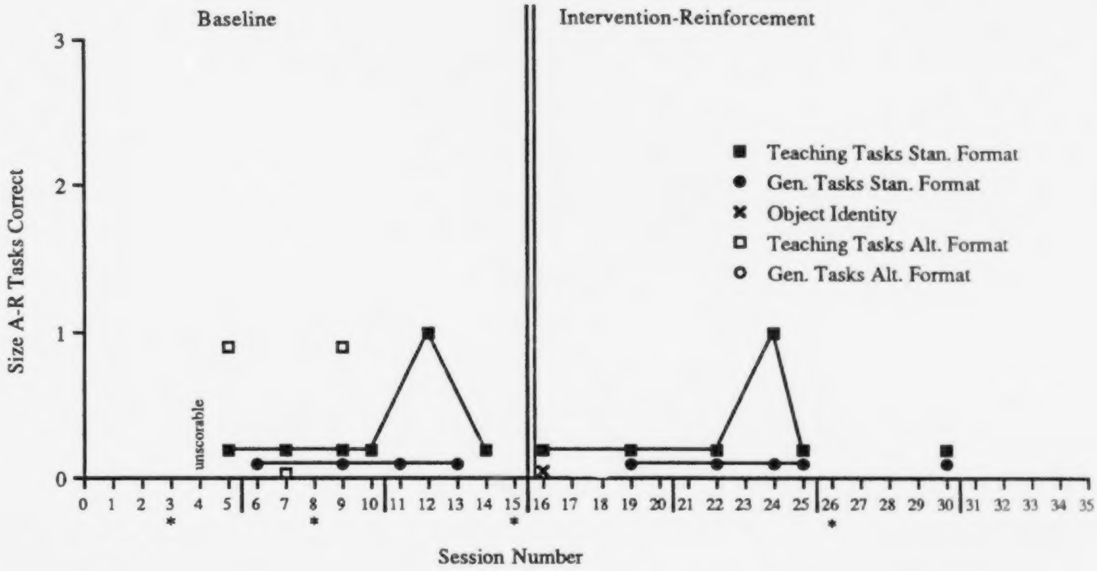


Figure 4.15. Brian—Number of correct responses on tests of Size tasks across conditions.

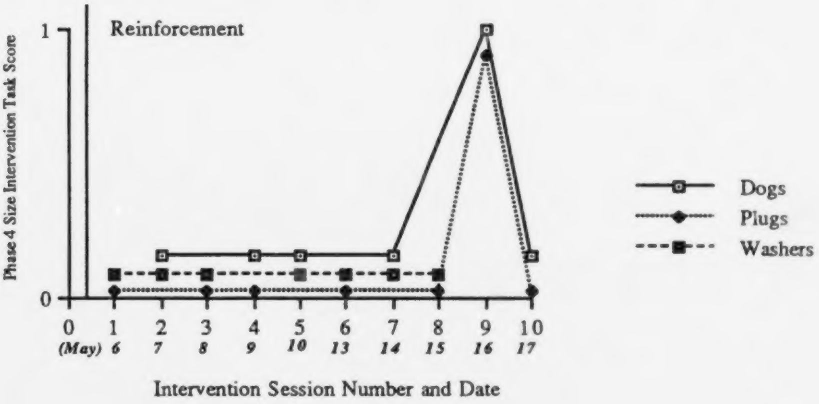


Figure 4.16. Brian—Phase 4 intervention on Size tasks.

concept "bigger." Of the 11 trials using the different pairs of objects, Brian made a mistake on only one trial. The "bigger" tasks were then administered during the study only when the validity of Brian's responses was suspect. Thus, in Session 4, after Brian's perseverative responses on the Size tasks were detected, the "bigger" tasks were administered, also yielding perseverative responses. On the subsequent 22 trials of the "bigger" tasks in the remainder of the study, Brian correctly identified the larger object of a given pair 100% of the time.

As seen in Table 4.12 and Figure 4.15, Brian was able to complete two tasks of the six that were administered using the alternate question format (33% success rate). The use of the alternate format, then, made no appreciable difference in Brian's performance on the Size A-R tasks.

Table 4.12

Brian—Data Breakdown of Size Sessions When Alternate Question Format Used

		Standard Format		Alternate Format	
Session Number	Task	Pass	Fail	Pass	Fail
5	Washers				√
	Dogs		unscorable		
	Erasers (SG)			√	
6	Plugs		√		
	S rings (SG)				unscorable
7	Plugs				√
	Dogs		√		
	Washers				√
9	S rings (SG)		√		
	Washers		√		
	Plugs		√		
	Dogs			√	
	Dogs		√		

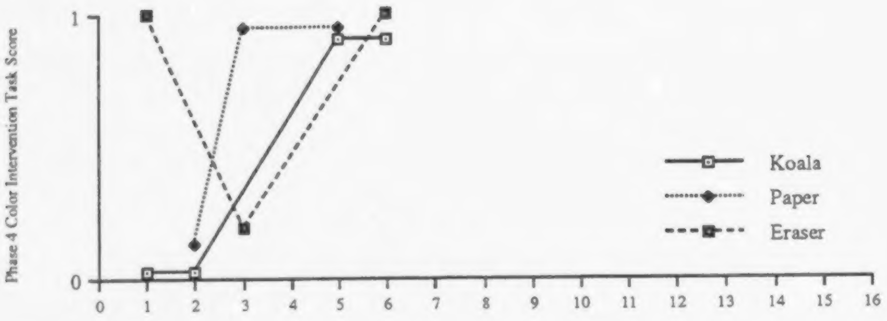
Brian was able to complete only one Size A-R task in one session after instruction was introduced (Session 24) despite the use of secondary reinforcement (introduced with the first intervention session). However, this specific task was correct only because Brian looked under the plexiglass before answering the Reality question. When Brian answered the Reality question with the plexiglass over the objects, phenomenist errors were made each time on both the instructional tasks and generalization tasks.

As seen in Figure 4.16 Brian was also able to complete Phase 4 of the instruction procedure correctly on only one occasion (intervention Session 9). With the exception of Session 9, Brian's answers during all phases of instruction were inconsistent, even when the investigator modelled the correct answer (as in Phase 2), or pointed to the correct answer after asking an Appearance or Reality question following Direct Instruction procedures. That is, regardless of the investigator's help, Brian often pointed to the apparently larger object when asked the Reality question, or to the objectively larger object when asked the Appearance question.

Although Brian was not able to complete Phase 4 correctly in intervention Sessions 7 and 8, in these sessions, and in Session 9, it was evident that Brian possessed some understanding of how the plexiglass changed the apparent size of the smaller object, and that its apparent size was not, in fact, the true size of the object. For example, in Phase 4 on the Plugs task in intervention Session 8, Brian correctly placed the plexiglass over the objects and pointed to the apparently larger object when asked to make the smaller object "look" bigger. When the investigator pointed to the apparently larger object and said, "Tell me about this one. It...," "Look bigger," finished Brian. When the investigator asked the Reality question, Brian removed the plexiglass and pointed to the objectively larger object. However, when the investigator replaced the plexiglass over the objects and asked the Reality question, Brian wavered between the objects and then picked the apparently larger object. This same pattern of responding was evident in all phases of Session 9, the difference being that Brian was able to answer the Reality questions correctly in Phase 4 when the investigator prevented him from removing the plexiglass. However, when tested on the tasks in this session (Session 24 in Figure 4.15), Brian was able to complete only one task correctly. It seems that the visual salience of the apparent size of the objects when the plexiglass was covering them dominated Brian's responses, and although he had some idea as to the A-R distinction, he was unable to consistently inhibit his response to the visually dominant larger object. This salience hypothesis is discussed more thoroughly in Chapter Five.

Intrasubject comparison of Color and Size tasks. Comparison of the Color and Size intervention sessions (Figure 4.17), demonstrates that instruction and mastery, in one task

Color Tasks



Size Tasks

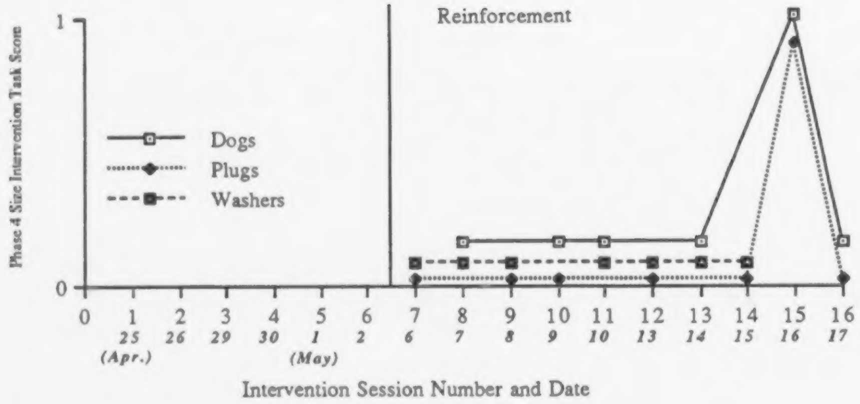
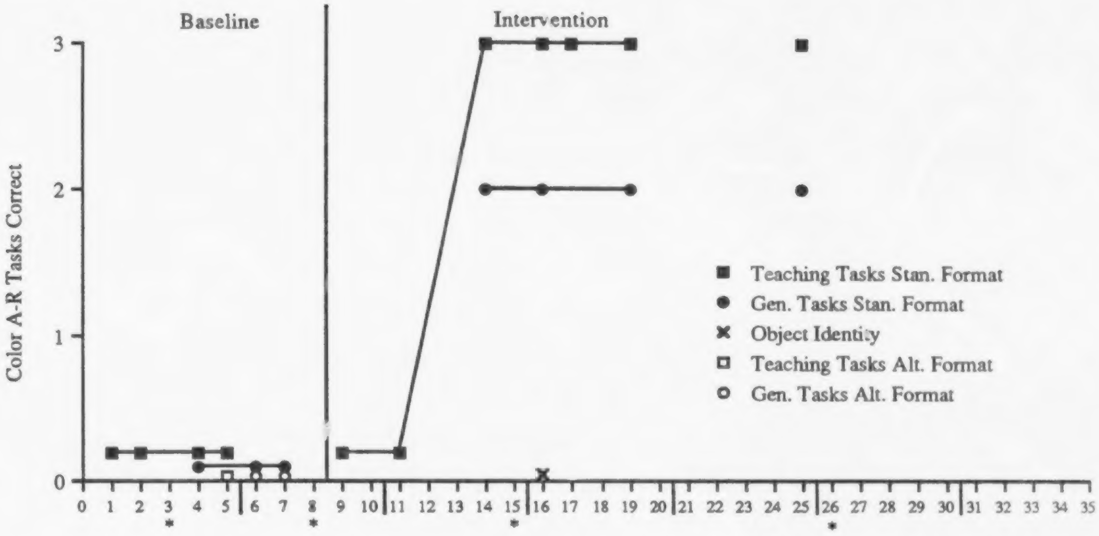


Figure 4.17. Brian—Comparison of Phase 4 intervention on Color and Size tasks.

Color Tasks



Size Tasks

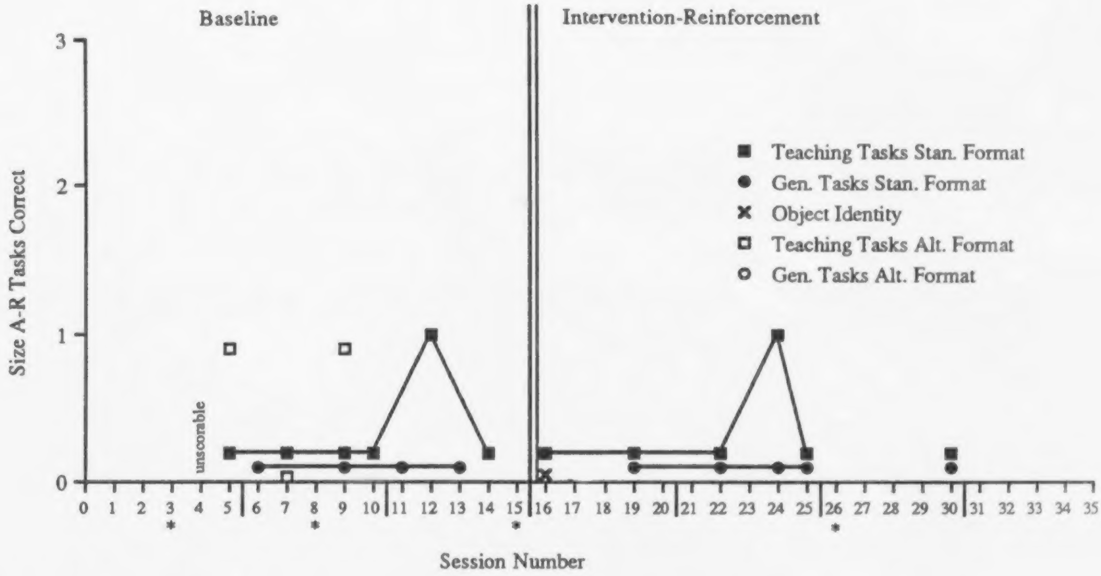


Figure 4.18. Brian—Comparison of performance on tests of Color and Size tasks.

domain (Color) had no effect on Brian's ability to achieve mastery in the other task domain (Size). As seen in Figure 4.17, Brian was able to complete Phase 4 of the instructional procedure on the Color tasks by intervention Session 5 without the use of secondary reinforcement, while he remained unable to complete Phase 4 on the Size tasks (with one exception) even with the use of secondary reinforcement. Similarly, when the Color and Size tasks were tested (Figure 4.18), criterion was achieved and maintained on the Color tasks by Session 14, with no effect on Brian's performance on the Size tasks. Thus, Brian did not generalize his understanding of the A-R distinction concept from the Color tasks to the Size tasks. Although he was able to overcome the salience of the deceptive color of the objects in the Color tasks, and report the object's true color, it appeared that the two task domains were different enough (with the Size tasks being more difficult for Brian), that the salience of the apparently larger object (when covered with the plexiglass) dominated Brian's responses.

Object Identity tasks. Like Nicky, and the other subjects in the study, Brian was unable to answer the second control question in any of the Object Identity tasks. When the plastic orange was presented to Brian in the Object Identity pretest, Brian did identify it as an orange. When asked what it was made of, however (control question 2), Brian replied, "orange." The investigator then asked Brian if it was a real orange or a pretend orange, to which Brian replied, "pretend orange," picked up a pencil, and tapped the "orange." The investigator continued with the Appearance and Reality questions, the answers to both questions being "orange."

In Session 1 (Figure 4.18), the investigator repeated the Object Identity task, using the plastic egg to see if Brian was able to complete the task using a different object. When the "egg" was revealed to Brian the task proceeded as follows.

I: What is this?
B: It's an egg.
I: You can touch it.
B: (Picks the egg up).
I: What is it made of?
B: Egg.
I: Is this a real egg or a pretend egg?
B: (No response).
I: Is this a real egg or a pretend egg?
B: (No response but handles and bends the egg).
I: (Pauses to allow Brian to discover the "egg's" properties) Is this a real egg or a pretend egg?

B: Pretend egg.

I: And what do you think it's made of?

B: Egg.

I: What is *this* egg made of? What is this *pretend* egg made of?

B: (No response).

I: (Replaces egg on mat) When you first saw the egg over here like this, what did you think it was?

B: Egg.

The investigator discontinued the task at this point because of Brian's obvious incomprehension.

In Session 2, however, the investigator attempted the task using a real orange and then repeated the task using the plastic orange. Brian was unable to answer the second control question on either of these tasks, and the tasks were discontinued.

The Rock Object Identity task was administered to Brian in Session 16, after criterion had been reached on the Color A-R tasks. However, not only was Brian unable to answer the second control question, he was also unable to identify the object as a "rock" in the first place. Hence, the administration of the task was discontinued.

When the Orange task was re-administered as a post-test, Brian again said "orange" when asked what it was made of. Contrary to the pretest, however, Brian said that the orange was a real orange when asked "Is this a pretend orange or is it a real orange?" It appeared that Brian's response to this question depended on the word order of the question. That is, if the investigator asked if an object was pretend or real, Brian replied that it was real (e.g., Orange post-test). If, on the other hand, Brian was asked if the object was real or pretend, he replied that it was pretend (e.g., Orange pretest). Although the investigator cannot be certain that this word order dependency accounted for Brian's answer when the Egg task was administered, the possibility seems likely given his answers on the Orange pre- and post-tests. However, because the second control question was not answered correctly in any of the Object Identity tasks that were administered, the results of the tasks cannot be interpreted.

Summary

In summary, an adequate level of trial reliability was obtained for Brian on both the Color and the Size appearance-reality distinction tasks, with percentage agreements ranging from 83.33% to 98.21%, and phi and kappa coefficients ranging from .89 to .96.

Instruction on the Color A-R tasks resulted in rapid acquisition, and only five instructional sessions were required before criterion was reached. Although Brian seemed to have some understanding that the plexiglass used in the Size A-R tasks altered the true size of the objects (judging from his tendency during many sessions to look under the glass before answering the Reality question) criterion was not reached on the Size tasks in spite of the use of edible reinforcement. Responding during the various phases of instruction on the Size tasks remained inconsistent. The investigator believed, however, that Brian may have been successful in this task domain if instruction had continued for a few more sessions.

Despite Brian's understanding of the A-R distinction as it applied to color, he did not generalize this understanding to the Object Identity tasks. Brian was unable to answer the second control question on any of the Object Identity tasks, and the administration of these tasks was discontinued.

Bruce

Description

Bruce is another subject of a visible minority for whom English is a second language. The family's native language is spoken in the home, despite the fact that the family has lived in Canada at least since Bruce was a toddler.

Bruce's mother reported that at one year of age, Bruce experienced convulsions as a result of a high fever, and that prior to this incident his development appeared to be normal. Bruce began attending public school at five years of age. He changed schools in 1986 and was enrolled in the autism program at the new school. Assessment of Bruce by a local hospital in 1987, determined that Bruce met the diagnostic criteria for autism. Intellectual assessment using the LIPS, the Raven Progressive Matrices Test, and the Merrill Palmer Scale, found that Bruce was functioning at the educable mentally handicapped to borderline range in nonverbal tasks. According to the most recent assessment, conducted in 1989 when Bruce's chronological age was 8.8, he was found to be functioning in the trainable mentally handicapped range according to the Stanford-Binet Fourth Edition (a language based test). Age equivalencies on the subtests of Stanford-Binet ranged from 2.7 to 4.3. Bruce achieved a nonverbal mental age of 5.2 on the LIPS, and scored in the EMH range on the Merrill Palmer test. It was noted that Bruce's greatest difficulty on this test was in answering "what" and "why" questions.

Bruce was 10.9 years of age at the beginning of the study and although enrolled at his school, he was integrated into a regular grade 3-4 classroom about 95% of the time.

On the LIPS administered by the investigator for the current study, Bruce achieved a nonverbal mental age equivalent of 5.1 (CA=10.9). Bruce's receptive vocabulary age-equivalent, as measured by the PPVT-R was 3.8, and his MLU on a language sample was 2.5.

Although cooperative during the study, Bruce's attention often needed to be re-directed to the task at hand. He had a tendency to look around the room or at his hands. When this occurred, the investigator provided a gestural and/or a verbal prompt (e.g., the investigator pointed to the display, or snapped her fingers and said, "Look Bruce. Watch what I'm doing."). Bruce was responsive to these prompts, and returned his attention to the task being presented accordingly. However, some doubt remains as to either how carefully Bruce was listening to the investigator, or how well he understood the questions when the various tasks were presented. For example, during the first administration of the Size baseline tasks, the investigator mistakenly asked Bruce, "What color looks bigger?" (instead of "Which one is bigger?" or "Point to the one that's bigger.") on two of the tasks. The investigator noticed her mistake when administering the second task. Bruce, however, showed no reaction to the error, and continued to point to one of the objects. This, along with Bruce's tendency to begin pointing to one of the objects before a question was even asked, led the investigator to believe that Bruce either did not really understand the task requirements at this point in the study, or that he simply was not listening carefully to what the investigator was saying.

Inter-Rater Reliability

Color tasks. As seen in Table 4.13, on a total of 61 observations of the Color tasks, no disagreements on either behavior occurrence or nonoccurrence were recorded. There were 50 agreements of behavior occurrence and 11 agreements of behavior nonoccurrence. Thus, total percent agreement, effective percent agreement on occurrence and effective percent agreement on nonoccurrence was 100, and the phi and kappa coefficients were 1.0.

Size tasks. As seen in Table 4.14, the inter-rater reliability check for Bruce on the Size tasks also produced no disagreements on behavior occurrence or nonoccurrence on the 55 observations. There were 22 agreements of behavior occurrence, and 33 agreements of behavior nonoccurrence. Total percent agreement, effective percent agreement on occurrence, and effective percent agreement on nonoccurrence were thus all 100, while phi and kappa correlation coefficients were both 1.0.

Table 4.13

Bruce—Inter-Rater Reliability on Color Tasks

		Observer 2		
		0	1	
Observer 1	1	A 0	B 50	A + B = 50
	0	C 11	D 0	C + D = 11
		A + C = 11	B + D = 50	N = 61

Table 4.14

Bruce—Inter-Rater Reliability on Size Tasks

		Observer 2		
		0	1	
Observer 1	1	A 0	B 22	A + B = 22
	0	C 33	D 0	C + D = 33
		A + C = 33	B + D = 22	N = 55

Results

Color tasks. As seen in Figure 4.19, repetition alone of the baseline Color tasks appeared to have resulted in Bruce's acquisition of the appearance-reality distinction concept as applied to color. However, as will be seen, this interpretation may not be justifiable.

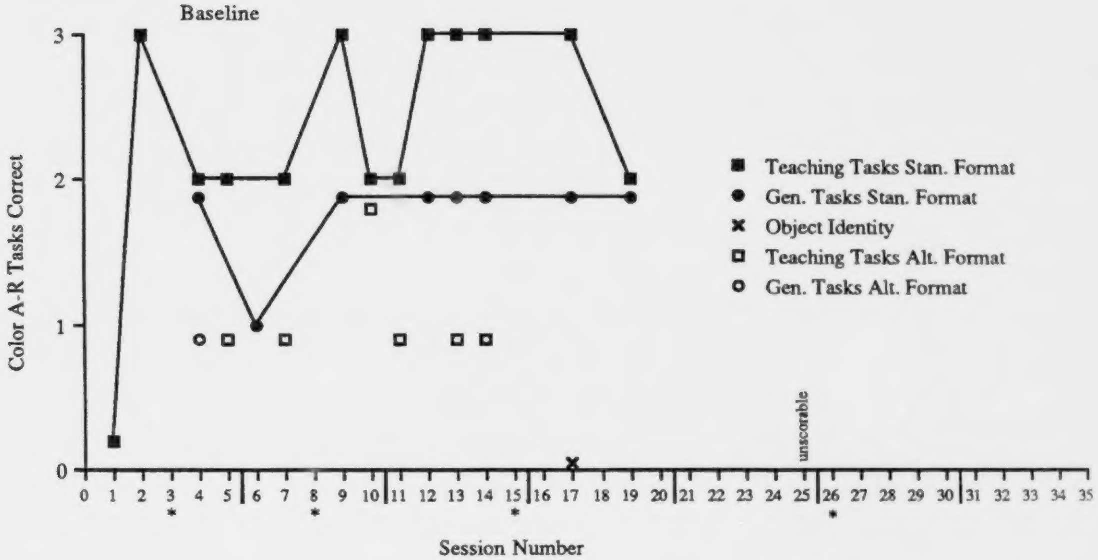


Figure 4.19. Bruce—Number of correct responses on tests of Color tasks across conditions.

Because baseline testing ultimately resulted in correct responding, the intervention phase of the study was never introduced to Bruce for the Color tasks. Although responding was inconsistent for the first 11 sessions, consistent correct responding was established in Session 12, and continued until Session 18. The inconsistent responding noted in sessions 1 through 11, and in Sessions 19 and 25 was attributable to the order in which the Appearance and Reality questions were asked.

On 12 out of 12 Color A-R items administered subsequent to Session 4 on which the Reality question was asked first , Bruce gave incorrect answers. Specifically, when Bruce was asked what color the object was in reality, he responded by stating the apparent color of the object, and when the Appearance question was subsequently asked, Bruce reported the true color of the object concerned. Conversely, when Bruce was asked the Appearance question first, he consistently answered both the Appearance and the Reality questions correctly on all but two trials aside from Session 1 (phenomenist errors were made on all trials of Session 1). The most logical interpretation of Bruce's insistence on reporting the apparent color first regardless of the question, is a language-based one. Because the apparent color of the object, when covered by the filter, was the most salient aspect of the situation for Bruce, this was reported first. The mere presence of the second question may have implied to Bruce that an alternative answer was required. Because the object at hand had originally appeared as a different color (e.g., the apparently green paper was white when originally presented), the original color of the object was the most logical alternative answer. Thus, the possibility exists that, because of language comprehension difficulties, Bruce did not fully understand the intent of the questions. Bruce provided an alternative answer because the investigator was apparently not satisfied with the first answer.

Alternatively, Bruce may simply have become accustomed to the order in which the questions were asked. It was not until Session 5 that the Reality question was asked first. Until this session, the Appearance question was always asked first. It may have been that Bruce intuitively understood the concept, but did not listen to the questions as they were being asked. Once a pattern had been established (i.e., Appearance first, Reality second), Bruce was not willing to change the order of his answers.

Session 25, the maintenance session, was unscorable because it became apparent that Bruce was alternating his answers regardless of the question asked. Specifically, when the investigator asked the same question repeatedly, Bruce consistently alternated his responses. The discovery of this pattern of responding makes interpretation of Bruce's performance on the Color tasks difficult. Because the investigator did not ask the same question more than once prior to this session, it is likely that Bruce was alternating his answers on the Color tasks throughout the entire study. Responding to the appearance of the object first, Bruce always reported the apparent color before reporting the true color.

Examination of Table 4.15 shows that Bruce was able to answer the A-R questions correctly when the alternate question format was used 100% of the time (9 out of 9 trials). Again, however, because the Appearance question was asked first in this format, Bruce was probably alternating his answers here as well. Unfortunately, the investigator did not explore this possibility by asking the same question more than once sequentially.

Size tasks. The data for Bruce's performance on the Size appearance-reality distinction tasks is presented in Figure 4.20. Examination of this figure shows that a stable baseline was obtained by Session 10, with Bruce consistently making phenomenist errors on the tasks. The introduction of instruction did not immediately result in task acquisition. However, by Session 23, (the 12th intervention session and the 8th session after secondary reinforcement was introduced), Bruce was able to complete the Size A-R tasks correctly. As seen in this figure, the tasks presented in Sessions 4 and 5 of the baseline, and those presented in Session 22 were unscorable. It was evident in these sessions that Bruce was alternating his answers, rather than responding consistently to a particular question that was asked. However, contrary to his performance on the Color tasks, Bruce did not alternate his

Table 4.15

Bruce—Data Breakdown of Color Sessions When Alternate Question Format Used

		Standard Format		Alternate Format	
Session Number	Task	Pass	Fail	Pass	Fail
4	Paper	√			
	Koala		√		
	Chalk (CG)			√	
	Crayon (CG)	√			
	Eraser	√			
5	Koala		√		
	Paper			√	
	Eraser	√			
7	Paper			√	
	Eraser		√		
	Koala	√			
10	Koala	√			
	Paper			√	
	Eraser		√		
	Eraser			√	
11	Eraser			√	
	Paper		√		
	Koala	√			
13	Paper			√	
	Eraser	√			
	Crayon (CG)	√			
14	Chalk (CG)			√	
	Crayon (CG)		√		
	Eraser	√			
	Paper	√			
	Crayon (CG)	√			
	Koala			√	

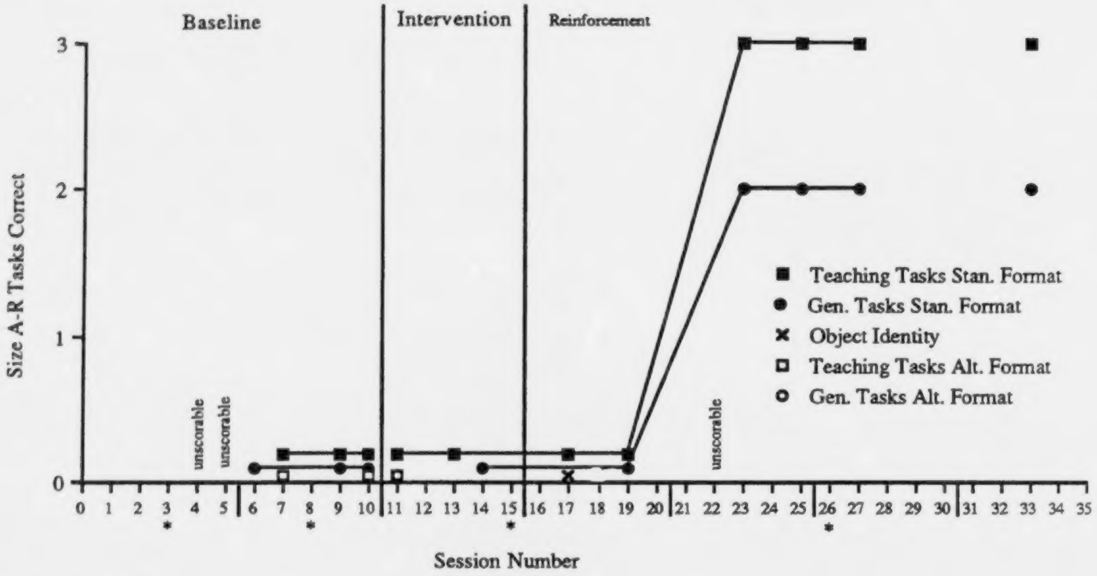


Figure 4.20. Bruce—Number of correct responses on tests of Size tasks across conditions.

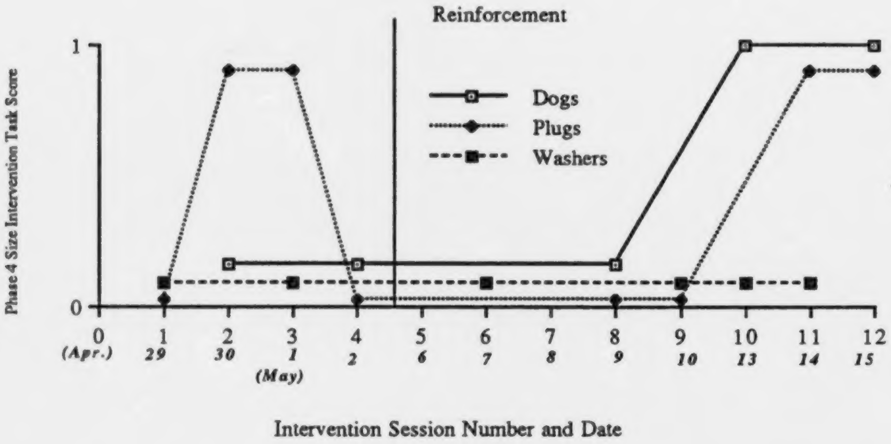


Figure 4.21. Bruce—Phase 4 intervention on Size tasks.

answers in the remainder of the sessions during either baseline or intervention, regardless of the question being asked in the sessions.

Because of the unscorability of the first two sessions of the Size task (Sessions 4 and 5), Bruce's understanding of the concept "bigger" was questionable. Therefore the "bigger" tasks were administered in Sessions 5, 6, and 7 to verify his grasp of the concept. Of seven trials administered in Session 5, Bruce identified the bigger object correctly on only three occasions. After the first error, the investigator prompted Bruce to listen, since sometimes the investigator was going to ask the *same* question. This prompt, however, did not result in consistently correct answers. During the second session in which the "bigger" tasks were administered (Session 6), the investigator praised Bruce when he gave the correct answer on the first trial, and explained further. "Sometimes I'm going to try to fool you, so you have to listen carefully." On the next trial, when Bruce was starting to point to the incorrect choice, the investigator stopped him and said, "No, listen to the question. Point to the *big* one." When Bruce changed his answer and pointed to the correct object, the investigator praised his choice saying, "Right. This is the big one." On the remaining 13 trials of the task in this session, and on all trials administered in Session 7, Bruce correctly identified the larger object regardless of what the object pair was, or on which side the larger object was located. By the end of Session 7 the investigator was convinced of Bruce's understanding of the concept "bigger." As seen in Table 4.16 (and in Figure 4.20), the alternate question format did not help Bruce answer the Appearance and Reality questions correctly. Of the 4 items administered using this format, Bruce was unable to pass any.

Figure 4.21 presents the data from Phase 4 of the intervention procedure. Because he was absent for Sessions 5 and 7, 12 intervention sessions were conducted with Bruce. Throughout the testing and instructional sessions on the Size tasks, Bruce consistently responded to the Appearance and Reality questions by saying "here" whenever he pointed to one of the objects. However, on two occasions in which the acetate was used in intervention Session 2, Bruce traced the lines on the acetate after the investigator had placed it over the objects, and said "here" without pointing to either object. Thus, Bruce's use of the word "here" did not necessarily indicate an informed answer. The use of the acetate was discontinued in Phase 3 of the first task presented in intervention Session 2.

As seen in Figure 4.21, Bruce was able to complete Phase 4 of the Plugs task in intervention Sessions 2 and 3, despite his inability to pass any of the Size tasks when they were tested at the beginning of the third intervention session (Session 13 in Figure 4.20). Although it appeared that Bruce understood the concept after intensive instruction involving

Table 4.16

Bruce—Data Breakdown of Size Sessions When Alternate Question Format Used

		Standard Format		Alternate Format	
Session Number	Task	Pass	Fail	Pass	Fail
7	Dogs		√		
	Washers				√
	Plugs				√
10	Dogs		√		
	S rings (SG)				√
	Erasers (SG)		√		
	Washers		√		
	Plugs				√

corrective and modelling procedures had been provided, it is possible that his correct responses on the Plugs task in these early sessions were due to chance. This interpretation is the most likely, since Bruce was unable to complete Phase 4 correctly on any task again, prior to intervention Session 10.

Because of Bruce's difficulty with the preliminary phases of instruction, Phase 4 of particular instructional sessions was not even presented for five of the tasks in the course of the study. Similarly, because of the difficulty that Bruce often experienced when Phase 4 was presented, in two instances instructional sessions concluded with an earlier phase (intervention Sessions 1 and 8). Despite these variations of the instructional procedure, the data are still presented in Figure 4.21 to graphically represent Bruce's failure to understand the appearance-reality distinction concept.

As seen in Figure 4.21, the introduction of reinforcement during instruction had no effect on Bruce's learning of the A-R concept. It took six more sessions after the introduction of reinforcement before Bruce demonstrated his understanding of the Size tasks. The investigator concluded, therefore, that Bruce's acquisition of the concept was primarily due to increased understanding of the task, rather than from increased motivation contingent on the provision of secondary reinforcement.

Two particular behaviors were noted in Bruce's responses to Phases 2 and 3 in intervention Sessions 1 through 11. The first was Bruce's insistence on selecting the apparently larger object when asked the Reality question, despite an immediately preceding model of the correct answer. For example, in Phase 2 of Session 9, (after the entire sequence had been modelled in Phase 1 in this session and in 8 previous sessions) the investigator modelled the correct answer to the Reality question with the plexiglass covering the objects.

> I: This one is *really* bigger (pointing to objectively larger object). Which one's *really* bigger Bruce?
>
> B: Here (pointing to phenomenally larger object)
>
> I: This one *looks* bigger (points to phenomenally larger object) but this one (points to objectively larger object) is *really* bigger (removes plexiglass while still pointing to objectively larger object). See? (Replaces glass) Which one's really bigger? This one (points to objectively larger object). Which one's *really* bigger?
>
> B: Here (Points vaguely at objectively larger object so quickly that the investigator is unsure of the object to which Bruce was referring).
>
> I: Point to which one is *really* bigger.
>
> B: (Points to phenomenally larger object).

A similar pattern of responding was evident throughout the instructional phase of the study on all Size tasks.

The second behavior that was observed with regularity throughout instruction, was Bruce's tendency to alternate his answer to the Reality question when the plexiglass was over the objects. When the plexiglass was not over the objects, however, Bruce could identify the larger object consistently and correctly. The use of the plexiglass confused Bruce, and resulted in indecision when answering the Reality question.

The investigator found that the time delay procedure (in which the correct answer was initially provided before Bruce had a chance to make an incorrect answer), and gestural cues (where the investigator either pointed to the correct answer, or indicated the correct answer with a gestural cue to assist Bruce in answering correctly) were necessary until Session 10, at which point Bruce began to demonstrate understanding of the task. During all phases of instruction prior to this session, Bruce's answers to the Appearance and Reality questions were inconsistent.

To conclude, the investigator believed that Bruce learned the requirements of the task in the course of instruction, since he was ultimately able to complete the A-R tasks correctly and consistently regardless of the order of the Appearance and Reality questions. This performance is in contrast to his performance on the Color tasks, where Bruce alternated his answers, and always responded to the apparent color of the objects before considering their true color.

Intrasubject comparison on the Color and Size tasks. Upon examination of Figure 4.22, it appears that although understanding of the Color A-R concept was demonstrated by Bruce before intervention was introduced, this apparent understanding did not generalize to the Size A-R tasks. However, because of the investigator's previously mentioned reservations regarding Bruce's understanding of the Color tasks (e.g., alternating of answers), a direct comparison of his learning in the two task domains is not possible.

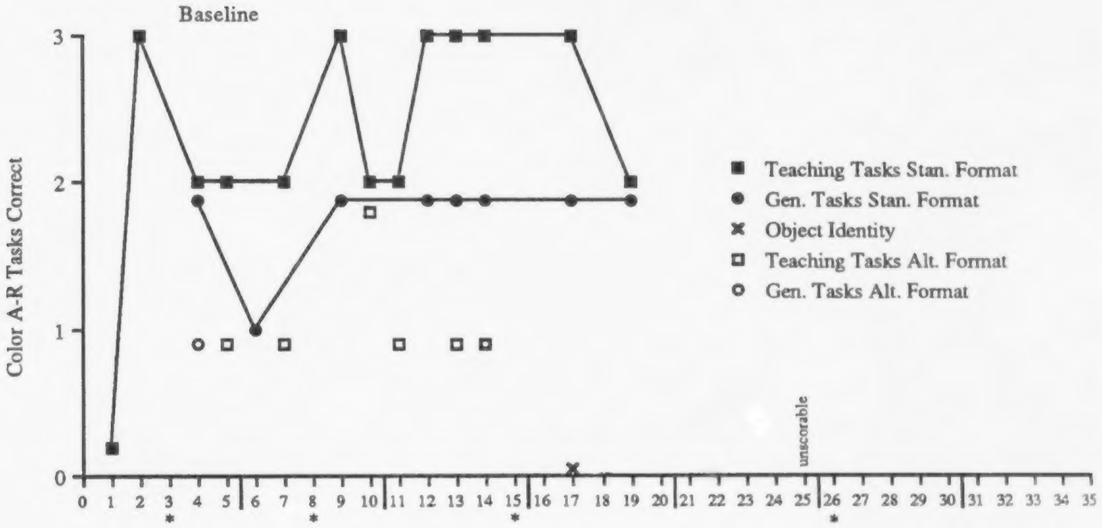
Object Identity tasks. The Object Identity tasks proved to be very difficult for Bruce to comprehend. For example, when the pretest task, using the "orange," was presented to Bruce, he spontaneously identified the object as an orange. After he picked it up he said, "It's a fruit." The investigator asked the second control question, "What is it made of?," to which Bruce replied, "lemonade." The investigator then asked, "Is this a real orange?," and Bruce replied, "yes." The investigator then inquired further, "Is it a real orange or a pretend orange?" and Bruce said that it was a pretend orange. However, when the investigator asked, "Can we get juice from *this* orange?," Bruce said "yes." Despite the inconsistencies in Bruce's responses, the investigator continued with the task to find out how Bruce would respond to the remainder of the questions. Hence, the investigator replaced the orange on the mat, and said, "When you saw this over here like this, what did you think it was?" Bruce replied, "yellow." The investigator tried again. "What did you think it was? What does it look like?" Bruce replied, "It's lemonade."

Bruce's responses during the administration of this task may have demonstrated his incomprehension of the task's requirements. Because the linguistic demands of the task seemed to be above Bruce's level of comprehension, the investigator modelled the Object Identity task using the plastic egg, providing the correct answers for the entire task.

> (Investigator exposes egg) What is this? It's an egg. I wonder what it's made of? I'm going to find out what it's made of. I'm going to pick it up. Oh! It's plastic! (Bruce echoed, "It's plastic.") Uh huh. It bends. It's all put together. It's plastic. (Investigator replaces egg on mat) Now, when I saw it over here like this, I thought it was an egg. It *looked* like an egg. But *really*, it's plastic.

The investigator thought that such a model might demonstrate the expectations of the task to Bruce more clearly. It was believed that the explicit discussion about what attributes made the egg a "fake" egg would assist Bruce in completing the other Object Identity tasks. This model was also similar to the pretraining models used by Flavell et al. (1986) and Flavell et al. (1987). Despite the inclusion of the model, Bruce remained unable to complete the Object

Color Tasks



Size Tasks

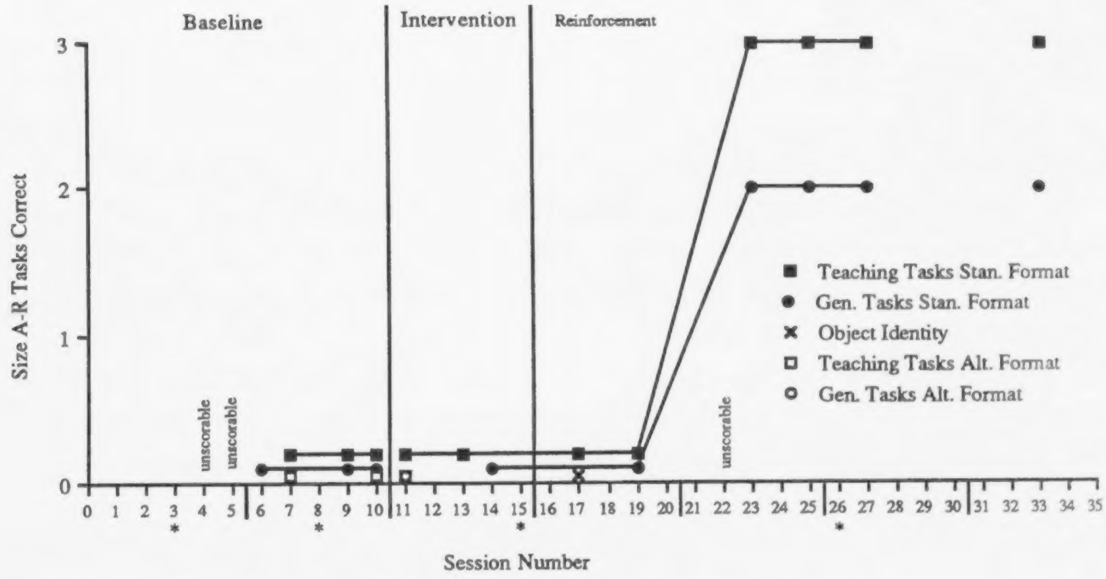


Figure 4.22. Bruce—Comparison of performance on tests of Color and Size tasks.

Identity task when the artificial flowers were used, and made similar errors on this and the Orange task when it was re-administered in Session 1 and Session 2 of baseline.

The investigator also attempted to make the artificial nature of the fake orange more obvious to Bruce by conducting the task with a real orange and subsequently re-administering the task using the fake orange.

I: What's this? (Indicates real orange)
B: Orange
I: (Giving Bruce the orange) What is it made of?
B: Orange.
I: Could you eat this orange? (Indicating orange)
B: No. At lunchtime.
I: OK. But could you eat this at lunchtime?
B: It's fruit.
I: (Removes real orange and repeats task with fake orange) What is this?
B: Orange.
I: (Gives "orange" to Bruce) What's this one made of? (Indicates orange)
B: Peel.
I: Could you eat this orange?
B: No.
I: Why not?
B: 'Cause it's lunchtime.

As seen in this excerpt, the utilization of the real orange failed to make the Object Identity task any easier for Bruce to understand.

Despite Bruce's inability to complete the Object Identity tasks during the pretest, the Rock generalization task was administered after criterion had been reached on the Color tasks in the interest of determining whether the task had become any easier for Bruce once another task domain was understood. Although criterion was eventually reached on the Color and Size tasks, Bruce was unable to generalize his understanding of the appearance-reality concept to the Object Identity tasks. When the Rock task was administered, Bruce was not able to verbalize its artificial nature. Before the investigator revealed the "rock" to Bruce, she instructed him to look at the object without touching it. However, once the rock was revealed Bruce immediately touched it. The investigator quickly removed his hand from the rock and said, "Just look at it. Tell me what it is," to which Bruce replied, "rock." Thus, even after he had touched it, he still apparently believed that the object was, in fact, a rock. The investigator then instructed Bruce to pick up the rock. Bruce did so , spontaneously said, "rock,"and put it down. Since the investigator was not convinced that Bruce had discovered the rock's artificial nature, she asked him to feel the rock. Bruce then picked it up again and squeezed it. While he was holding the rock, the investigator asked, "Is this a pretend rock or a real rock?" to which Bruce replied, "a real rock." The investigator attempted to further clarify Bruce's understanding of the properties of the rock in question.

> I: If you had *this* rock and threw it at someone would it hurt them? Or if you threw it at a window, would *this* rock break a window?
>
> B: Yeah.
>
> I: Would *this* one?
>
> B: No.
>
> I: Why not?
>
> B: (mumbled)

The investigator's repeated questioning regarding the rock's capability of breaking a window (e.g., "Would *this* one?") implied that a different answer was required. Although Bruce noticed this syntactic clue and changed his answer, he was unable to verbalize the reason for his answer. Thus, it is likely that he changed his answer simply in response to the repeated question, rather than as a consequence of knowing and understanding the correct answer.

Bruce's inability to understand the artificial nature of the objects used in the Object Identity tasks, and thus to complete the tasks, was further seen in the Object Identity post-test. Upon re-administration of the Orange task, the following exchange took place.

> I: I'm going to show you something. I don't want you to touch it, but just look at it. (Reveals "orange") What is it?
>
> B: (Touches orange) Orange.
>
> I: Now you can pick it up.
>
> B: Peel.
>
> I: Is this orange a pretend orange or a real orange?
>
> B: Real orange.
>
> I: Could you peel *this* orange? (touches orange)
>
> B: Yes.
>
> I: Would this one be good to eat?
>
> B: Lunchtime.

The task was discontinued at this point since Bruce was unable to answer even the control questions. The investigator observed that when Bruce was presented with any of the Object Identity tasks and asked whether an object was "pretend or real" (or "real or pretend"), Bruce consistently responded that the object was real, if the word order of the enquiry was "pretend or real", and that the object was a pretend one when the word order was reversed.

In other words, Bruce's responses to this question (similar to Brian's responses), were contingent on the word order used by the investigator. Doubt is raised, therefore, as to Bruce's comprehension of the question.

An assessment conducted in 1989 indicated that Bruce had particular difficulty in answering "what" and "why" questions. Bruce also had difficulty in answering "what" and "why" questions in the Object Identity tasks of the present study. When the investigator asked why the "orange" couldn't be eaten or what one of the Object Identity task items was made of, Bruce either did not respond, or made a response that did not answer the question (e.g., when asked what the artificial flowers were made of, Bruce responded, "red flowers").

To conclude, Bruce's performance on all of the Object Identity tasks suggested that the level of language comprehension required to complete the tasks as designed in the present study, was beyond Bruce's capabilities. Therefore, conclusions regarding Bruce's capability to generalize the appearance-reality distinction concept to object identity cannot be made.

Summary

To summarize, inter-rater reliability calculations for Bruce on both the Color and the Size tasks resulted in complete agreement between the observers in both task domains. Thus all percentage agreements were 100 and all phi and kappa coefficients were 1.0.

Bruce demonstrated apparent understanding of the Color A-R tasks with repetition of the baseline tasks alone, and thus instruction was never introduced to Bruce for the Color tasks. Consistent correct responding was established by Session 12. However, the investigator found that Bruce was responding to the Appearance and Reality questions in a very prescribed manner. Regardless of the question asked first, Bruce consistently responded to the apparent color of the object first, and then responded to the true color of the object. The investigator concluded that because Bruce's true understanding of the appearance-reality distinction was suspect, interpretation of the results of this task domain was not possible.

A stable baseline on the Size tasks was established by Session 10, with Bruce giving phenomenist answers to all of the tasks. After intervention was introduced, Bruce reached criterion on the Size tasks by the twelfth instructional session. The introduction of secondary reinforcement had no apparent effect on Bruce's demonstration of understanding of the task.

Bruce was unable to complete any of the Object Identity tasks correctly at any point in the study. He apparently did not understand the requirements of the task since, like the previously discussed subjects, he was unable to answer the control questions correctly.

Steve

Description

Steve's chronological age was 12.1 at the beginning of the study. In 1987, the year in which Steve and his mother moved to the province, the local school board assessed Steve using the Autism Behavior Checklist and the Psycho-Educational profile, and found that he met the criteria for the category of severe behavior disorder—autism. Steve was enrolled in the autism program at a local public school, a program in which he still participates.

Steve's mother reported that Steve's developmental delay was first noticed when he was two years of age. In 1989, Steve's mother completed Rimland's E2 Form (an autism diagnostic checklist), and sent it to Rimland's Autism Research Institute for interpretation. The subsequent report stated that Steve achieved a score of +31, identifying him as a "classical" case of autism, which puts Steve in the Kanner Autism subgroup. According to Rimland, only about 10 percent of autistic children score as high as +20.

On the tests used in the current study, Steve achieved a nonverbal mental age-equivalent of 5.4 (CA=12.1), an MLU of 2.6, and a receptive vocabulary age-equivalent of approximately 3.5 on the PPVT-R. Because of Steve's low raw score on the PPVT-R, no exact age equivalent is available. Although Steve's receptive vocabulary age-equivalent was below the initial inclusion criteria, he was retained as a subject in the interest of seeing whether it would be possible to teach the appearance-reality distinction concept to such a low-scoring student.

Steve's behavior during the study was marked by frequent outbursts of laughing, guttural noises, rocking, hand and forearm biting, and clapping. The laughing behavior and the noises that Steve made were especially prevalent in the preliminary stages of the study, and were probably attention-seeking behaviors. At first the investigator turned away from Steve and ignored this behavior, re-engaging attention by praising Steve after he had been quiet for a few seconds. Although this procedure may have been effective in the long-term, the investigator found that a sharp "no!" was more effective in eliminating the behavior quickly.

Steve's hand and forearm biting was most frequent after instruction had begun, and seemed to be a reaction to frustration encountered when the investigator corrected his errors during the various phases of instruction. Steve would put his face close to the investigator's, attempt to make eye contact with the investigator, and bite his own hand. When this occurred, the investigator avoided eye contact, asked Steve to put his hands down, and concluded the session, after ensuring that Steve had been successful on one of the tasks.

Inter-Rater Reliability

Color tasks. The inter-rater reliability data for Steve's Color tasks is presented in Table 4.17. Out of 35 scorable observations, there were no disagreements between the observers on occurrence or nonoccurrence of behavior. The observers agreed that there were no instances of behavior occurrence, and 35 instances of behavior nonoccurrence. Total percent agreement, effective percent agreement on occurrence of behavior, and effective percent agreement on nonoccurrence of behavior were thus all 100 while phi and kappa correlation coefficients were both 1.0.

Table 4.17

Steve—Inter-Rater Reliability on Color Tasks

		Observer 2		
		0	1	
Observer 1	1	A 0	B 0	A + B = 0
	0	C 35	D 0	C + D = 35
		A + C = 35	B + D = 0	N = 35

Size tasks. In Table 4.18 the inter-rater reliability data reveals that there was one disagreement between the observers on the Size tasks for Steve out of 43 observations. Specifically, the investigator recorded a nonoccurrence of behavior, while the second observer recorded an occurrence (Cell D). There were nine agreements of behavior occurrence and 33 agreements of behavior nonoccurrence. As a result, the total percent agreement was 97.67, the effective percent agreement on occurrences was 90, and the effective percent agreement on nonoccurrences was 97.06. The phi correlation coefficient was .93 and the kappa coefficient was .94.

Table 4.18

Steve—Inter-Rater Reliability on Size Tasks

		Observer 2		
		0	1	
Observer 1	1	A 0	B 9	A + B = 9
	0	C 33	D 1	C + D = 34
		A + C = 33	B + D = 10	N = 43

Results

Color tasks. The results of Steve's performance on the Color appearance-reality distinction tasks are presented in Figure 4.23. Throughout the baseline period Steve made phenomenist errors on each task regardless of the question order. Steve's pattern of responses on the Color tasks indicated that he was not alternating his answers, as might have been implied by the presence of a second question. That is, despite the presence of a second question (i.e, the Reality question), Steve did not change his answer, but reported that the true color of the object was the same as its apparent color. Neither does the data suggest that Steve's responses on the Color tasks were perseverative, since he consistently, and correctly, identified the true color of each object before the filter was placed over the object (which consequently changed the object's color). Hence, the investigator believed that Steve's data was a valid indication of his understanding of the appearance-reality distinction concept as it pertained to the property of color, at least as it was assessed in the current study.

As seen in Table 4.19 and Figure 4.23, the alternate question format did not enhance Steve's ability to complete the Color A-R tasks correctly. Of the eight trials in which the alternate format was used, Steve completed the A-R tasks incorrectly on all of them, again giving phenomenist answers.

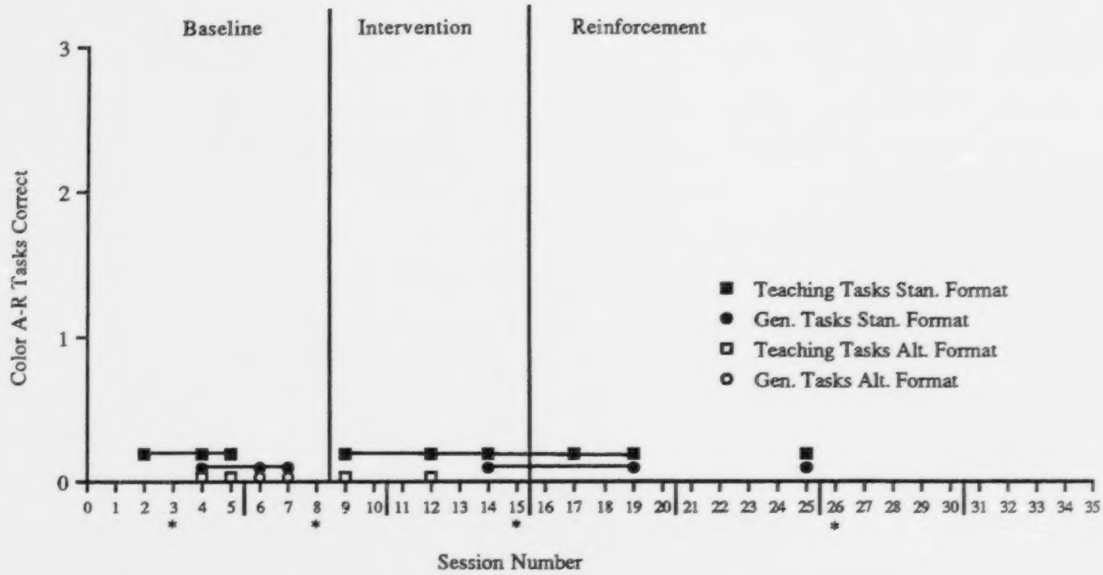


Figure 4.23. Steve—Number of correct responses on tests of Color tasks across conditions.

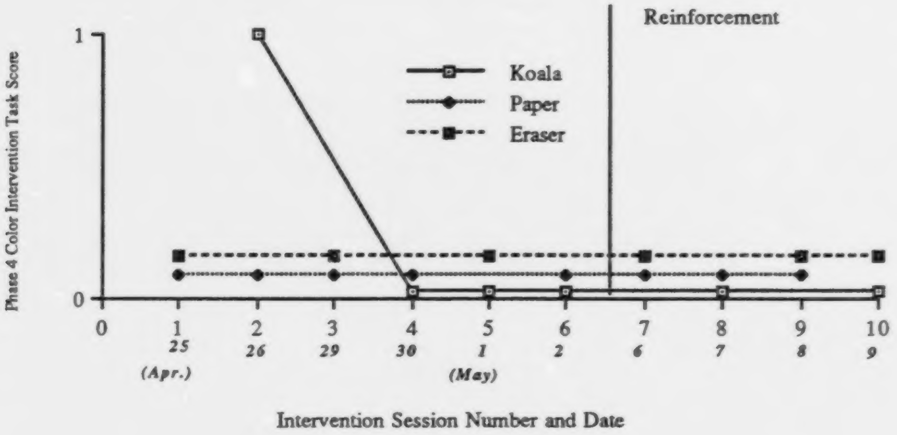


Figure 4.24. Steve—Phase 4 intervention on Color tasks.

Table 4.19

Steve—Data Breakdown of Color Sessions When Alternate Question Format Used

		Standard Format		Alternate Format	
Session Number	Task	Pass	Fail	Pass	Fail
4	Koala		√		
	Chalk (CG)		√		
	Eraser				√
	Crayon (CG)		√		
	Paper		√		
5	Koala				√
	Paper		√		
	Eraser				√
6	Crayon (CG)				√
	Chalk (CG)		√		
7	Chalk (CG)				√
	Crayon (CG)				√
9	Eraser		√		
	Paper				√
	Koala		√		
12	Paper		√		
	Eraser				√
	Koala		√		

As illustrated in Figure 4.24, Steve was unable to demonstrate understanding of the Color appearance-reality distinction tasks during Phase 4 of the instructional procedure on all but one task, during one session (intervention Session 2) for the duration of the study. The investigator believed, however, that Steve's success on this particular administration of Phase 4 of the Koala task was due to chance, primarily because two previous administrations of Phase 4 of this task during the same session yielded incorrect responses to the Reality question. In addition, Steve's failure to complete Phase 4 of any subsequently presented task confirmed his lack of understanding of the concept.

Steve's most notable behaviors during the Color task instruction, once he became familiar with the tasks, was his tendency to either verbalize each step with the investigator when it was being modelled in Phases 1 and 2, to echo each step after the investigator had verbalized it, or to complete a sentence (correctly) when the investigator paused in the middle of a modelled presentation. For example, in Phase 1 of the Eraser task in intervention Session 7, the following exchange took place.

I: (Partially covers eraser) It's *really* pink.
S: It's really pink.
I: But it *looks* blue.
S: But it looks blue.
I: (Covers eraser completely with filter) It *looks* blue but it's *really* pink (removes filter).
S: It's really pink.
I: (Completely covers eraser) It *looks* blue...
S: (Not looking at display) And it's really pink.

Similarly, during Phase 4 of the Eraser task in Session 6, after Steve had placed the filter over the eraser the investigator asked, "What did you do Steve?" Steve said, "You made the *paper* look green [emphasis added]," a repetition of what the investigator had said to Steve on previous tasks in previous sessions. The investigator believed that Steve, like Nicky in early sessions, although able to learn the linguistic syntax involved in the tasks without much difficulty, did not understand the semantic or pragmatic demands of the tasks.

Various behaviors that Steve exhibited during instruction on the Color tasks were similar to those of the other subjects. One such behavior was Steve's habit of answering the Reality question correctly, as long as a portion of the object's true color was visible. At the point where the object's true color was completely obscured by the filter, Steve changed his answer to the Reality question, and reported the apparent color of the object instead. Like Bruce during instruction of the Size tasks, Steve too would answer the Reality question incorrectly immediately after providing a correct answer that had been reinforced, or after a model of the correct answer had been presented. Hence, after 10 intervention sessions, Steve's responses during the instructional phases were no more consistent than they had been during the first session.

Although secondary reinforcement was introduced in intervention Session 7, it had no effect on Steve's pattern of responding. Steve continued to make the same kinds of errors as he did before reinforcement was introduced. The investigator concluded, therefore, that a lack of motivation did not account for Steve's inability to complete the tasks.

In conclusion, despite the use of secondary reinforcement, and 10 instructional sessions, Steve was unable to reach criterion on the Color tasks. The investigator believed that Steve's relatively low level of language expression and comprehension, compared to the other subjects in this study, was the primary cause of his inability to demonstrate understanding of the tasks.

Size tasks. As can be seen in Figure 4.25, Steve's baseline for the Size tasks was much less stable than that for the Color tasks. This lack of stability was thought to be a consequence of Steve's tendency to alternate his answers regardless of the question asked. Although the investigator attempted to control for alternating responses by asking the same question repeatedly, complete control may not have been achieved.

As in the Color tasks, the use of the alternate question format did not make an appreciable difference in Steve's ability to answer the Size A-R tasks correctly (Figure 4.25 and Table 4.20). Of six trials in which this format was employed, Steve passed a task on only one occasion in Session 7 (14% success rate).

The "bigger" tasks were administered on five occasions in the course of the study (in Sessions 2, 5, 7, 10, and 13), to ensure Steve's understanding of the concept "bigger." Understanding of the concept was especially important to verify, given Steve's tendency to alternate his answers even when asked the same question. Steve's understanding of the concept, however, was found to be secure. Of a total of 39 trials of the tasks administered in those five sessions, Steve was able to correctly select the larger object 100% of the time. Incomprehension of the concept, "bigger," was not therefore, a factor in Steve's inability to complete correctly the appearance-reality distinction tasks.

At no point after intervention was introduced (which was concurrent with the introduction of secondary reinforcement), was Steve able to complete the Size A-R tasks correctly. His responses remained either unscorable (due to alternation of responding) or phenomenist. The order in which the Appearance and Reality questions were asked also had no effect on either the scorability, or the type of error that Steve made on any of the tasks.

Figure 4.26 presents Steve's Phase 4 intervention data of the Size tasks. Although Steve completed Phase 4 correctly on two tasks in intervention Session 1, and on one task in intervention Session 4, the investigator attributed this success to chance, because of Steve's failure to complete this phase correctly on subsequent tasks and in subsequent sessions.

Similar to Bruce on the Size tasks, Steve required gestural prompts to answer the questions correctly during the various phases. Specifically, the investigator, after asking an Appearance or Reality question in Phases 2 or 3 of the Size tasks, needed to either point directly to the correct answer, or point in the general direction of the correct answer for Steve to be successful. Despite the repetition of the tasks over 10 instructional sessions, this

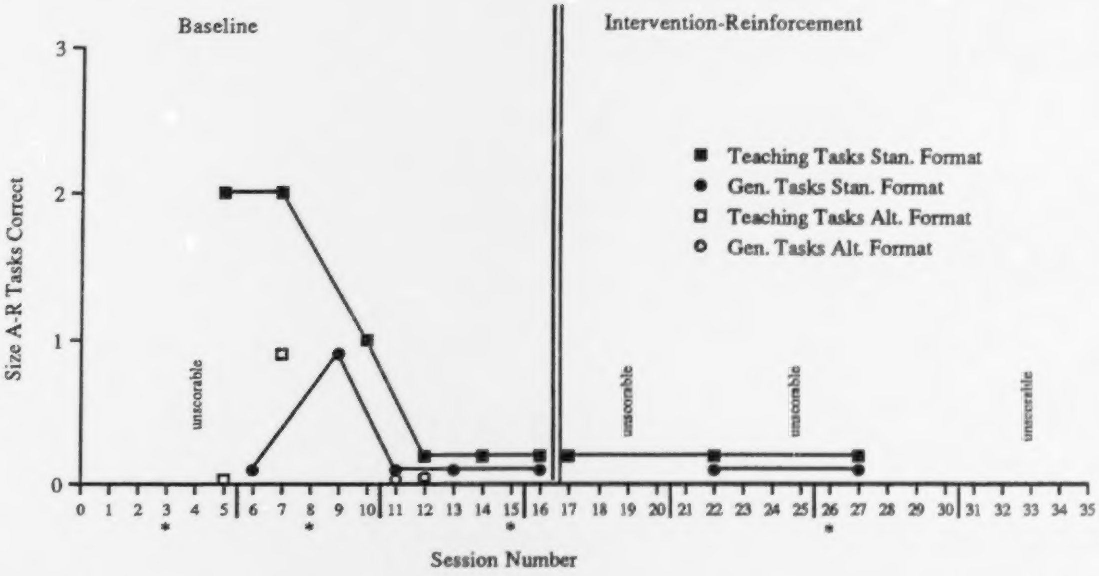


Figure 4.25. Steve—Number of correct responses on tests of Size tasks across conditions.

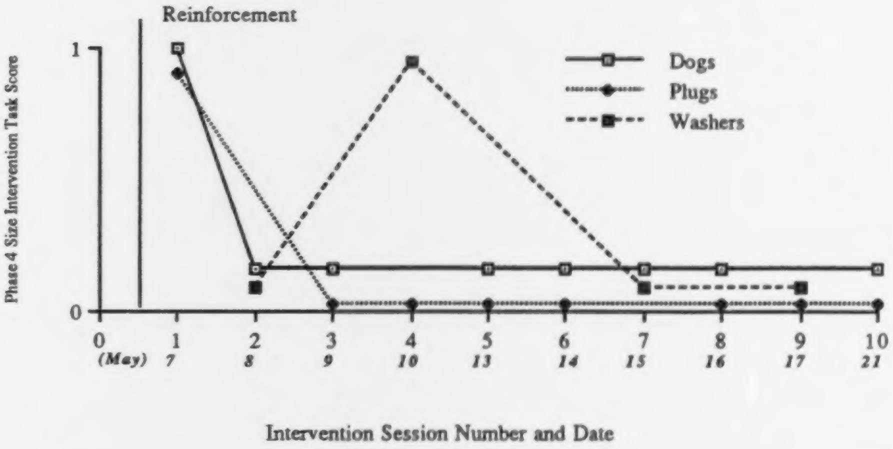


Figure 4.26. Steve—Phase 4 intervention on Size tasks.

Table 4.20

Steve—Data Breakdown of Size Sessions When Alternate Question Format Used

		Standard Format		Alternate Format	
Session Number	Task	Pass	Fail	Pass	Fail
5	Erasers (SG)		√		
	Dogs				√
	Washers		√		
7	Plugs				√
	Washers	√			
	Dogs			√	
11	Erasers (SG)		√		
	S rings (SG)				√
12	Washers				√
	Plugs		√		
	Dogs				√

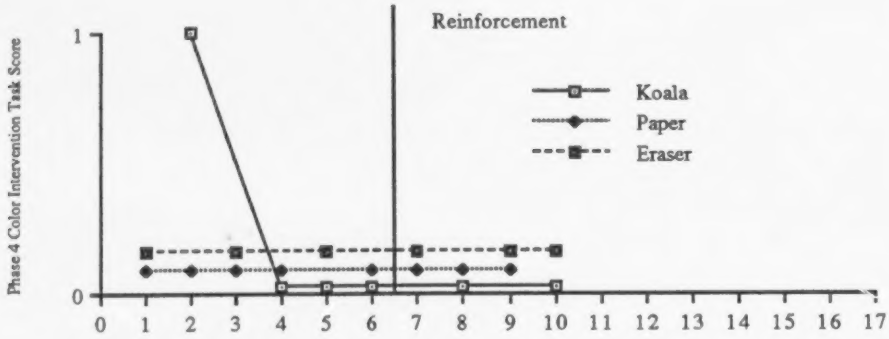
gestural prompt was necessary for the duration of the study. When the prompt was not provided, Steve usually incorrectly answered the Reality question in these phases.

Steve's difficulty with the Size tasks was also evident in his insistence that the apparently larger object was "really" the larger object, despite immediately preceding models of which object was "really" bigger, accompanied by demonstration of the same (removal of the plexiglass that showed which object was "really" larger). Although such models were provided throughout the study, they did not help Steve to select the correct answer to the Reality question.

In conclusion, Steve was also unable to reach criterion on the Size tasks despite the use of secondary reinforcement, models, and gestural prompts. Steve's behavior on these tasks indicated that although he understood the concept "bigger," he did not understand that although the plexiglass altered the apparent size of the objects, their true size remained unchanged.

Intrasubject comparison on the Color and Size tasks. When Steve's performance on the Color and Size tasks was compared (Figures 2.27 and 2.28), it was apparent that both tasks were equally difficult for Steve to comprehend, and criterion in either task domain

Color Tasks



Size Tasks

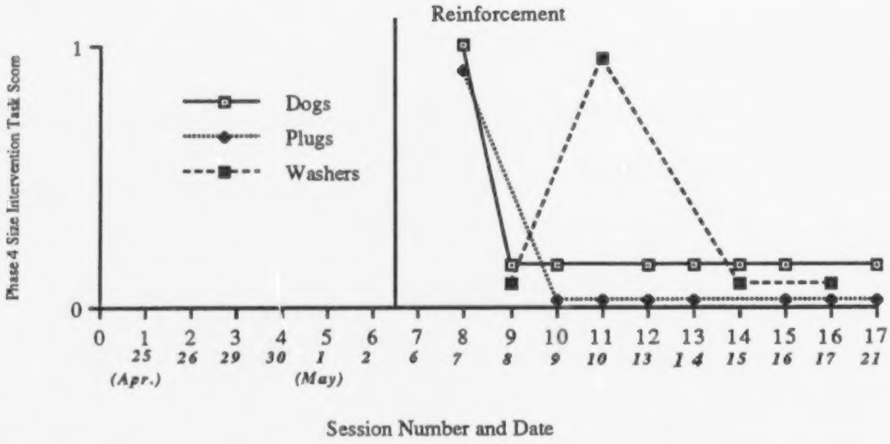
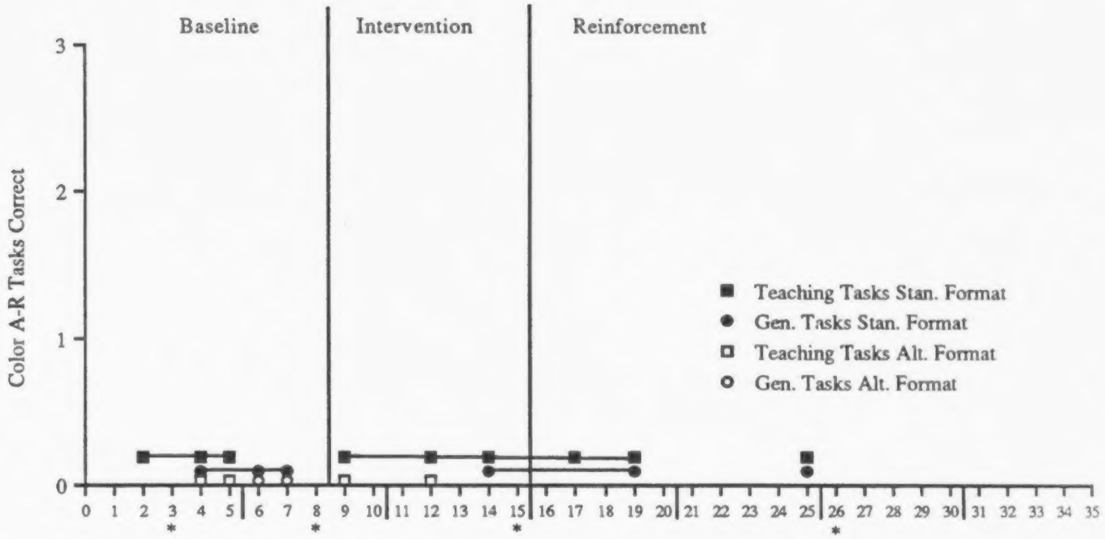


Figure 4.27. Steve—Comparison of Phase 4 intervention on Color and Size tasks.

Color Tasks



Size Tasks

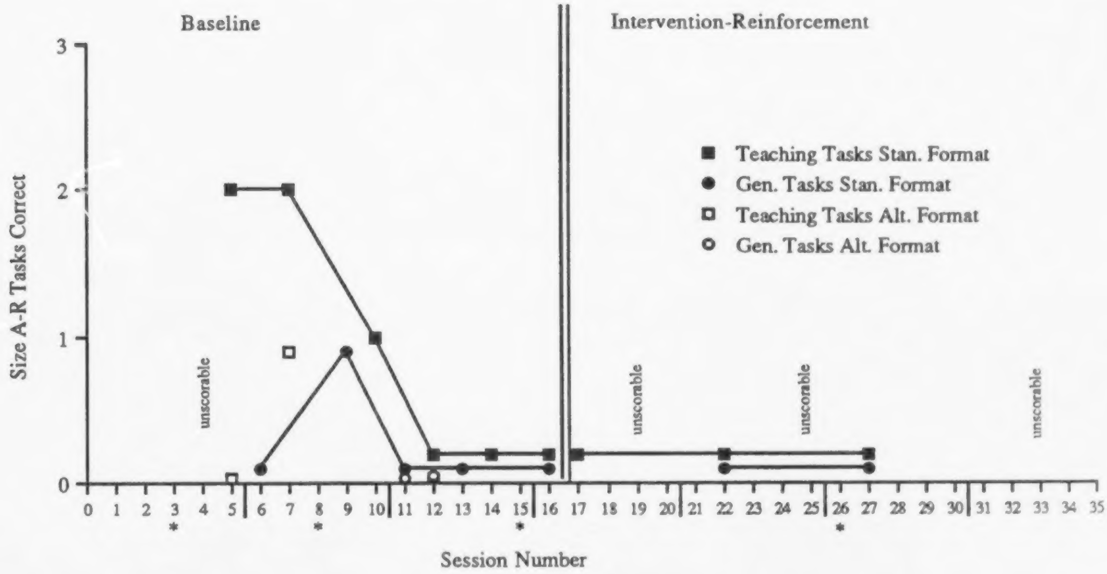


Figure 4.28. Steve—Comparison of performance on tests of Color and Size tasks.

could not be reached. As seen in Figure 2.28, reinforcement of the A-R concept through its instruction as applied to a different property did not help Steve in comprehending the concept in either task domain. That is, the introduction of instruction in the Size tasks, did not help Steve understand the A-R concept in the Color tasks.

The use of secondary reinforcement during instruction also had no effect on Steve's acquisition of the tasks in either domain. As seen in Figure 4.27, Steve's performance on Phase 4 of the Color tasks remained unchanged. Although there was some indication that Steve could complete Phase 4 correctly in the initial instructional session, the use of secondary reinforcement was not thought to have been responsible, particularly since correct responding was not maintained. As discussed above, Steve's correct completion of Phase 4 during the initial instructional session was thought to be due to chance.

Object Identity tasks. Similar to the other subjects in the study, Steve was also unable to understand the requirements of the Object Identity tasks. When the Orange task was administered during pretesting, Steve spontaneously picked up the "orange" after identifying it as an orange when it was first revealed to him. When the investigator asked Steve what it was made of, Steve replied, "an orange" while handling it and squeezing it. The investigator then asked, "Is it a real orange or is it a pretend orange?" Steve answered that it was a pretend orange, and proceeded to sniff it. Once the orange was replaced on the mat, the investigator asked the Appearance and Reality questions, with Steve's answer to both questions being, "an orange."

The Egg Object Identity task was modelled for Steve during the first baseline session in the same manner as it was modelled to Bruce, and for the same purpose (i.e., to draw Steve's attention to the particular attributes of the plastic egg that distinguished it from a genuine egg). However, as with Bruce, the modelling procedure did not help Steve in understanding the task. The administration of the Object Identity task using a real orange also did not help Steve complete the task. He remained unable to answer the control questions when the real orange was used, when the task was re-administered using the plastic orange, or when the artificial flowers were used. The administration of the Object Identity tasks was discontinued when the investigator failed to elicit responses to the questions and Steve began laughing and making noises.

In spite of Steve's failure to reach criterion on either the Color or the Size tasks, the Rock task was administered to Steve at the end of the study, to determine if the materials used would make a difference in Steve's comprehension of the task. When the "rock" was revealed to Steve, and the investigator asked him what the object was, Steve identified it as a rock. Steve picked the object up after he was instructed to do so. The manner in which Steve picked the rock up indicated that he found the rock to be unexpectedly lighter than anticipated, thus confirming to the investigator that Steve probably believed the object to be a real rock. Steve squeezed the "rock," replaced it on the table and poked it. When the investigator asked Steve what the "rock" was made of, Steve replied, "rock." The investigator then asked, "Is this rock a pretend rock or a real rock?" Steve replied that it was a real rock. Similarly, when Steve was presented with the plastic orange on the Object Identity post-test, he said that the object was made of "orange" when asked the second control question, and when asked if the orange was a pretend orange or a real orange, replied that it was "a real orange." Because Steve was unable to answer the control questions on the Object Identity tasks, interpretation of his performance is impossible.

Summary

In summary, inter-rater reliability measures yielded 100% agreement on the Color tasks and only slightly less than 100% agreement on the Size tasks when applied to Steve's data. An acceptable level of trial reliability was thus obtained.

Throughout the study, Steve made phenomenist errors on the baseline and the intervention measures on both the Color and Size task domains. The use of the alternate question format during baseline did not make a difference in Steve's responses to the Appearance and Reality questions. The introduction of tokens exchangeable for snacks also did not influence Steve's responding during task instruction.

Although Steve's understanding of the concept "bigger" was established, initial baseline data on the Size tasks was rather unstable. This was thought to be due to the investigator's unfamiliarity with Steve's method of responding. Once the investigator had determined Steve's pattern of responding, the investigator controlled for it by repeating questions to determine if a reliable response was being given. Hence, the investigator believed that the latter baseline data points were a more valid indication of Steve's understanding of the Size tasks. In addition, a total of four sessions of the Size tasks were deemed unscorable because of Steve's tendency to alternate his answers regardless of the question asked (e.g., when the Reality question was repeated, Steve alternated his answer each time).

Steve was also unable to understand the appearance-reality distinction as it was applied to the Object Identity tasks. On the investigator's attempts to elicit answers on the various questions, Steve either did not respond, was echolalic, or laughed and made noises. Although he was able to identify each object according to what it was representing (i.e., an orange, an egg, a rock, flowers), Steve was unable to answer the second control question (What is it made of?). Hence, Steve's responses to the Appearance and Reality questions on these tasks cannot be interpreted.

Summary of Results

Examination of the results of inter-rater reliability across subjects revealed that a high degree of agreement was obtained for all subjects on both the Color and the Size tasks. The minimum total percent agreement obtained was 91.18, with the maximum being 100. Effective percent agreement on behavior occurrences ranged from 75 to 100, while the range of effective percent agreement on behavior nonoccurrences was from 96.15 to 100. Phi coefficients ranged from .85 to 1.0 and kappa coefficients ranged from .84 to 100. These results enabled the investigator to conclude that the behaviors that were being measured were discrete enough to allow a second observer to attain the same scores as the investigator in most instances, and achieve trial reliability.

To summarize the results, of the five subjects included in the present study, three (Brian, David, and Nicky) demonstrated mastery of the Color appearance-reality distinction tasks after intervention. One subject (Bruce) appeared to gain mastery of the tasks simply through repetition of the baseline tasks, although because of methodological problems, this understanding was suspect. One subject (Steve) failed to achieve mastery on the Color A-R tasks, and the use of a secondary reinforcement procedure did not assist him in completing the Color A-R tasks correctly. Three subjects (Nicky, David, and Bruce) attained criterion on the Size A-R tasks as well. With the exception of Brian, the subjects who achieved criterion in one task domain also achieved criterion in the other task domain (Nicky, David, and Bruce). Steve alone was unable to achieve criterion on either task.

None of the subjects, however, were able to complete the Object Identity tasks at any point during the study. Because the investigator could not elicit an appropriate answer to the second control question in these tasks ("What is it made of?"), administration of the Object Identity tasks was either discontinued, or the results were uninterpretable. The investigator believed that the difficulty that the subjects had with the Object Identity tasks demonstrated the greater linguistic complexity of these tasks (at least when using the current presentation format) relative to the Color and Size tasks. If the linguistic requirements of the Object Identity tasks account for the subjects' failure on these tasks, it seems logical to assume that linguistic complexity also accounted for the various subjects' difficulties in the other task domains.

Flavell et al. (1986), and Leslie (1987, 1988) noted that the development of understanding of the appearance-reality distinction seems to occur simultaneously across domains. That is, mastery of property and identity A-R tasks occurs at the same age (roughly by four years of age) in normally developing children, when the tasks involve distinctions in color, size, or object identity. The performance of the subjects in the present study, however, suggests that for autistic children, mastery of the different task domains does not occur simultaneously. While the Size tasks appeared to be more difficult than the Color tasks for most of the subjects in the present study, none of the subjects were able to correctly complete any of the Object Identity tasks. In fact, none of the subjects were even able to answer all the control questions on the Object Identity tasks correctly. Thus, although the Object Identity tasks *may* have been of equal conceptual difficulty for the subjects, the linguistic difficulty posed by the format of the task made it too difficult for these subjects to comprehend, given their documented expressive and receptive language deficits.

CHAPTER FIVE

DISCUSSION

In this chapter, each subject's performance on the different appearance-reality tasks is discussed in relation to that of the other subjects. First, the post-test data on the Color, Size, and Object Identity tasks are considered. Next, each subject's performance on the Color and Size tasks is considered in relation to the other subjects, and patterns of behavior and errors are discussed. The chapter continues with an examination of the difficulties the subjects experienced with the Object Identity tasks and suggestions are made as to why this difficulty was encountered. Next, the effectiveness of the instructional methodology is considered and includes an evaluation of the Direct Instruction procedure, the method of task presentation, and the effectiveness of the reinforcement procedure and the question format used in this study. The penultimate section of the chapter notes the limitations of the current study and suggests ideas for future research that would clarify, confirm, and extend the present findings. The final section of the chapter summarizes the study's results.

Post-test Data

The subjects' post-test data is presented in Table 5.1. After intervention, two subjects (David and Bruce) were able to complete two of the post-test tasks correctly. Two subjects (Brian and Nicky) completed one post-test task correctly, and one subject (Steve) remained unable to complete any of the post-test tasks correctly after intervention was completed.

Examination of Table 5.1 shows that David and Bruce were able to complete the Color and Size post-tests correctly. That is, after instruction, these subjects were able to correctly identify that an object's apparent color or size after it had been distorted, was not the object's true color or size. Brian, however, answered only the Color Appearance and Reality post-test questions correctly, making a phenomenist error on the Size post-test (i.e., answering that the smaller object's phenomenal state after the plexiglass was put in place was also its absolute state). Nicky, on the other hand, answered the Size post-test questions correctly, and made a phenomenist error on the Color post-test (i.e., the apparent color of the object when distorted was the object's true color). Based on Nicky's consistently correct performance on all of the Color tasks after concept understanding had been obtained, however, the investigator believed that Nicky's mistake on the post-test was due to inattention. On the Color post-test Steve gave a phenomenist answer. Steve's answers on the Size post-test were unscorable. Because of the pattern of performance demonstrated by Steve during the study, the investigator asked the same question (i.e., either the Appearance

Table 5.1

Subject Post-Test Performance

				Type of Error		
Subject	Post-test task	Pass	Fail	Phenomenist	Realist	Other
David	Color	√				
	Size	√				
	Object Identity		√			√
Bruce	Color	√				
	Size	√				
	Object Identity		√		√	
Brian	Color	√				
	Size		√	√		
	Object Identity		√		√	
Nicky	Color		√	√		
	Size	√				
	Object Identity		√			√
Steve	Color		√	√		
	Size		unscorable			√
	Object Identity		√			√

or the Reality question) a number of times consecutively on the post-test, accompanied by a verbal prompt to listen carefully. This resulted in Steve simply alternating his answers regardless of the question that was asked or repeated.

Color and Size Appearance-Reality Task Performance

Although the study was not strictly a multiple baseline design, the Color and Size task data is presented in this format in Figures 5.1 and 5.2 respectively, for ease of comparison. Comparison of the five subjects' performance on the Color tasks (Figure 5.1) indicates that four of the subjects were able to reach criterion when tested on both the teaching tasks and the generalization tasks. Intervention was not introduced for one subject (Bruce), since he appeared to gain understanding of the concept through repetition of the tasks in a testing format. However, Bruce's apparent success was probably an artifact of the investigator's questioning technique (i.e., consistently asking the appearance question first), and thus Bruce's understanding of the concept was questionable. Of the remaining subjects who reached criterion on the Color tasks, mastery was demonstrated within 6, 8, and 3 intervention sessions respectively for Brian, Nicky and David.

In comparing the subjects' performance on the Size tasks (Figure 5.2), it can be seen that following intervention, three of the five subjects (Nicky, Bruce and David) demonstrated understanding of these tasks. Nicky demonstrated her understanding of the concept after 7 intervention sessions, Bruce demonstrated his understanding after 11 sessions, while David needed only 2 sessions before demonstrating his understanding.

The data of the present study suggest that the tasks representing the Size domain of the appearance-reality distinction were more difficult for most of the subjects to understand than were the tasks representing the Color domain. For example, although Brian could correctly complete Phase 4 of the Color tasks within five instructional sessions, he remained unable to complete Phase 4 on the Size tasks on all but one occasion. Although Bruce's acquisition of the concept within the Size domain cannot be directly compared to his acquisition of the concept within the Color domain because of methodological problems, 12 intervention sessions were required before Bruce could consistently complete Phase 4 of the Size domain tasks (Bruce was absent for 2 sessions). Similarly, although Nicky seemed to have at least a vague understanding of the A-R distinction with the Color tasks (seen in her inconsistent performance on Phase 4 of the Color tasks), she was unable to complete Phase 4 of any Size task until the seventh instructional session. The only subject for whom the two task domains seemed equivalent in difficulty was David. David was able to complete Phase 4 on both the Color and Size tasks by the third instructional session. The investigator believed that the observed differential difficulty between the two task domains was a function of the instructional procedure. This possibility is discussed in a later section of the chapter.

The subject who achieved criterion the most rapidly in both task domains of the present study (David), also scored the highest on the assessment of nonverbal mental age, and had the longest MLU. This finding is in line with those of Eisenmajer and Prior (1991) who examined the relationship between success on false-belief tasks and pragmatic language skills in autistic children. It took only two instructional sessions before David demonstrated understanding in both of these task domains when tested. This finding, however, was not consistent across subjects. For example, although Nicky was the second subject to achieve criterion on the Color and Size tasks, and had the second longest MLU, Brian scored higher than Nicky on the receptive vocabulary age-equivalent measure, and Steve scored

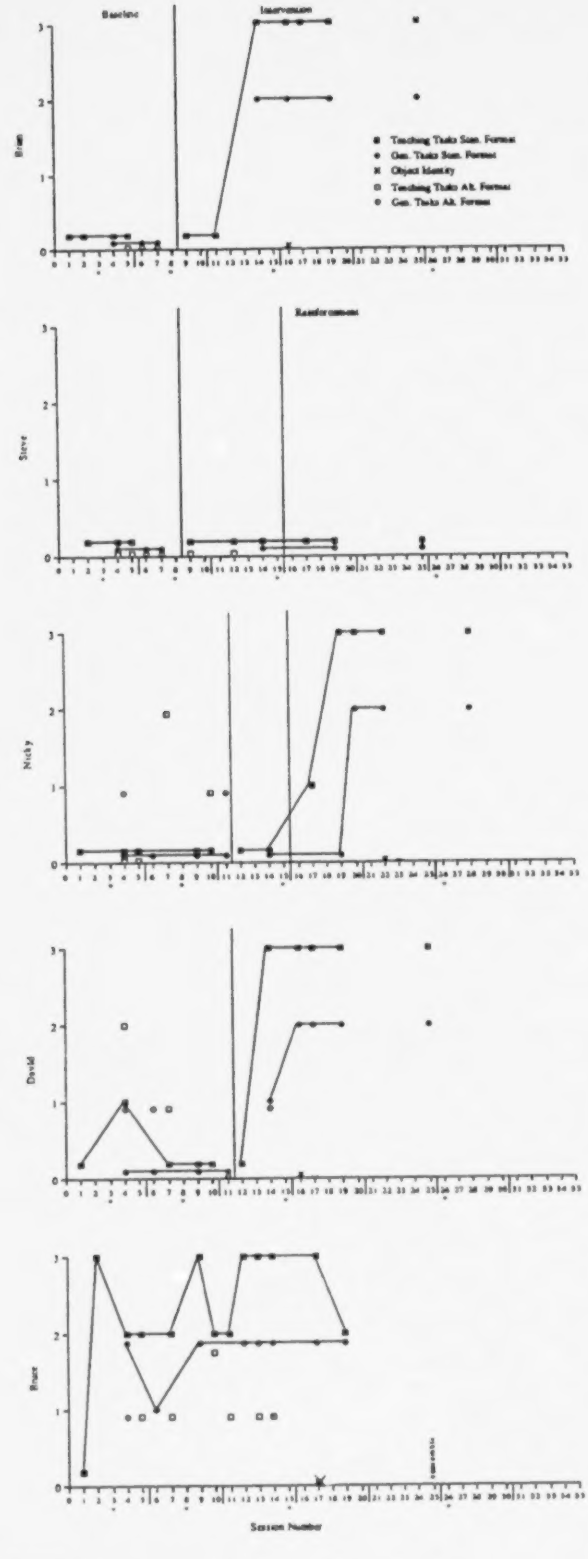


Figure 5.1. Intersubject comparison on Color tasks.

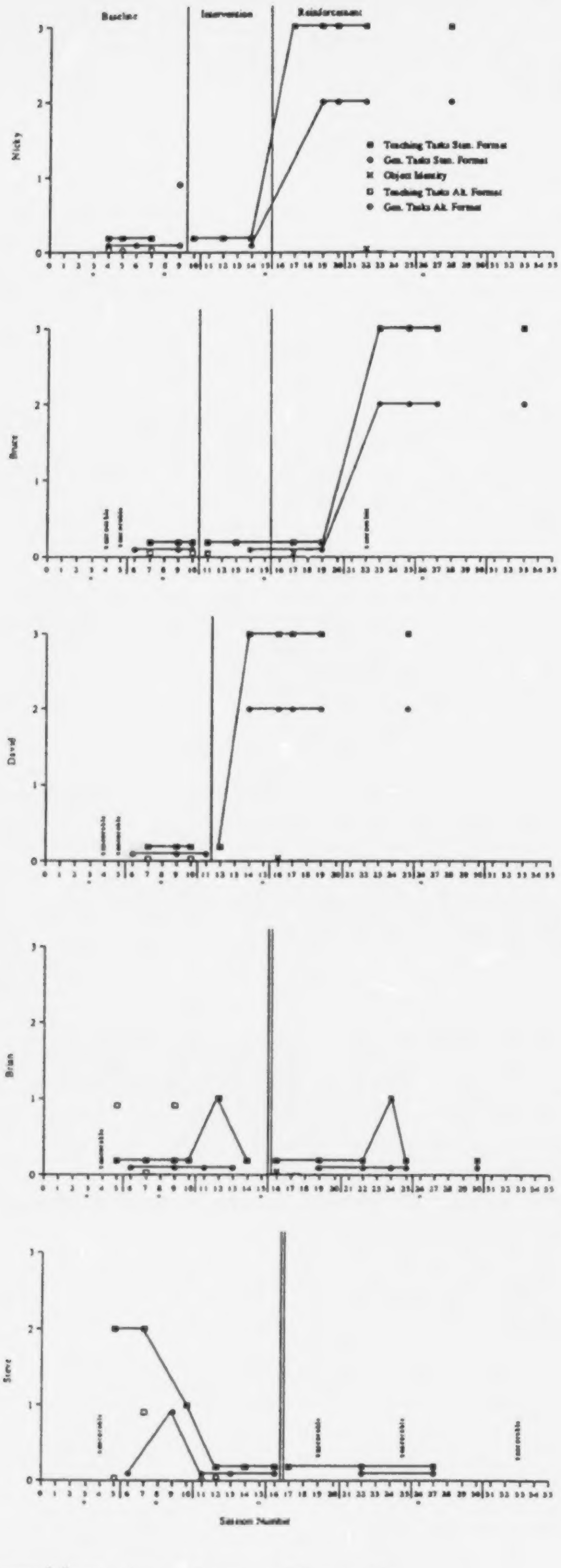


Figure 5.2. Intersubject comparison on Size tasks.

slightly higher than her on the nonverbal mental age measure (although when the standard error of measurement is taken into consideration, the difference is eliminated). Similarly, Steve (who did not achieve criterion in either task domain), had the second highest nonverbal mental age, the third highest MLU, and the lowest receptive vocabulary age-equivalent as measured by the current instruments.

Similar to Eisenmajer and Prior's (1991) findings, there was no relationship between chronological age and reaching criterion in the task domains in the current study. David's clear superiority over the other subjects in terms of MLU, receptive vocabulary, and nonverbal mental age, however, and the speed with which he attained criterion in the task domains suggested that the role of pragmatic language ability in completing theory of mind tasks needs to examined more systematically in future research.

Examination of Behavior and Error Patterns

At the beginning of the study, all of the subjects (with the exception of Bruce) made phenomenist errors on the Color tasks. That is, when asked the Reality question on these tasks, Brian, David, Steve, and Nicky reported that the apparent color of an object when covered by the filter, was the object's true color. Similarly, Nicky, David and Bruce consistently made phenomenist errors during baseline on the Size tasks. When his responses were scorable, Steve's answers too were phenomenist. These results were consistent with those of Flavell et al. (1983), Flavell et al. (1986), and Baron-Cohen (1989a). The subjects in these studies also had a strong tendency to give phenomenist answers in Color and Size appearance-reality distinction tasks. After instruction, however, Nicky, David and Brian (and perhaps Bruce to some extent) were able to overcome the salience of the illusory color of the stimulus, and report the true color of the object concerned.

During instruction on the Color tasks, all of the subjects enjoyed looking through the different colored filters and covering their hands. The investigator used these materials to try to reinforce the concept. It was thought that since the subjects were more familiar with the true color of their skin than with objects used in the tasks, the A-R concept might be more easily understood when their own hands were used. Hence, the investigator tried teaching the A-R concept using the filters and the subjects' hands. Although three of the subjects did ultimately reach criterion on the Color tasks, the effectiveness of this particular procedure cannot be isolated from other contributing factors.

A particularly interesting behavior observed in the course of the study, was the insistence of some of the subjects in selecting the apparent color or size in answer to the Reality question during instruction, despite the continual presence of the investigator's model, correction procedures, and the repetition of the Reality question. For example, during instruction on the Color tasks, Nicky and David answered the Reality question correctly when it was asked repeatedly while the investigator gradually covered the object with the filter. As soon as the object was completely covered, however, they changed their answer to the Reality question and reported the apparent color of the object, even though the instructional sequence would dictate that the same answer was expected. This insistence was not prolonged for Nicky and David, though, since criterion was achieved in 8 and 3 intervention sessions respectively.

The Size tasks, on the other hand, seemed to be more difficult for Steve, Bruce (during the initial instructional sessions), and Brian, and these subject's insistence on selecting the phenomenally larger object when asked the Reality question was more persistent. Three explanations could account for Steve's insistence on selecting the phenomenal color and size when asked the Reality question in both task domains, and the same insistence of Bruce and Brian on the Size tasks when asked the Reality question. First, the salience of the stimulus may have dominated these subjects' perceptions of the tasks. The Reality question was asked while the objects were in their phenomenal state. Even if Steve, Bruce and Brian understood the lexical terms "really" and "looks like," the visual evidence may have been too overwhelming to ignore. A similar explanation has been put forward by Flavell et al. (1983), in describing young normal children's performance on appearance-reality tasks, and by Russell, Mauthner, Sharpe and Tidswell (1991) in describing autistic children's performance on false-belief tasks.

Secondly, given the expressive and receptive language difficulties of the subjects, these boys may not have understood the distinction between the terms "really" and "looks like." For example, when the subjects were asked the Reality question in the Size A-R tasks, the word "bigger" (a concept that these subjects did understand), may have been the most prominent word, and the word to which these subjects responded. Thus, regardless of what the investigator modelled or indicated with gestural cues, when required to answer the question independently, Brian, Bruce, and Steve always chose the object that appeared larger. Consequently, sometimes the objectively larger object was selected in response to the Reality question (when the plexiglass was not over the objects), and sometimes the phenomenally larger object was selected (when the plexiglass was over the objects). This explanation would be consistent with Shatz's (in Owens, 1984) speculation that children with MLUs of approximately 1.5 to 2.75 may respond to a single action word within a request, and not to the entire form. Thus, Steve, Bruce, and Brian may have been responding only to the directive "point", and the adjective "bigger." In the Color tasks, Steve's incomprehension of the distinction between the terms "looks like" and "really" may have led him to base his answers solely on the visual salience of the apparent color of the objects when he was required to answer the Reality question independently (i.e., without a model or a prompt).

A third explanation of these results is consistent with the theory of mind interpretation That is, the subjects who failed to achieve criterion on either the Color or Size A-R tasks, had an autism-specific deficit in their ability to mentally represent the objects' apparent color or size, and the objects' true color or size simultaneously, a deficit that was unameliorable to instruction for some subjects. This explanation, however, appears to be the least likely of the three, given the subjects' language difficulties that have been noted, the linguistic complexity of the tasks, the short-term nature of the intervention, and the fact that Brian was able to achieve criterion in one of the task domains.

The investigator believed that a combination of the first two explanations accounted for Steve's, Bruce's and Brian's inability to demonstrate their understanding of the appearance-reality distinction concept in at least one task domain. It is possible that the salience of the stimuli in the task domains along *with* the subjects' difficulty in understanding the terms used by the investigator, combined to prevent these subjects from demonstrating understanding of the tasks. It must also be kept in mind that Steve and Bruce achieved the lowest receptive vocabulary age-equivalent scores on the PPVT-R compared to the other subjects (approximately 3.5, and 3.8 respectively), and, along with Brian, had the lowest MLU's. Although these subjects' poor receptive and expressive vocabulary as measured by these assessments does not conclusively prove that they are responsible for the subjects' apparent incomprehension of the tasks (particularly when the standard error of measurement of the test is taken into consideration), their language difficulties may be indicative of the subjects' task comprehension difficulties.

Despite the receptive language difficulties that might have contributed to the subjects' incomprehension of the task requirements, the subjects' verbalizations during task instruction were substantial. Specifically, once the subjects were familiar with the instructional procedure, all of them either verbalized the procedure with the investigator when the task was being presented, or completed the investigator's statement when a task was being modelled, and the investigator paused in the middle of the presentation. Nicky even imitated the instructional procedure, and "taught" the Color tasks to her imaginary friend. When these tasks were presented in a test format, however, the subjects were unable to answer the Appearance and Reality questions correctly, at least at the beginning of the study. Thus, although the subjects had the memory competencies to reproduce the syntactic forms, it was obvious that comprehension lagged behind. This behavioral observation strengthens the argument that the language used in the appearance-reality tasks played a major role in these autistic children's ability to demonstrate understanding of the concept.

Object Identity Task Performance

The Object Identity tasks appeared to be the most difficult for all of the subjects to understand. None of the subjects who demonstrated understanding of the appearance-reality distinction in at least one domain, were able to generalize this understanding to the Object Identity tasks. Although some of the subjects could correctly label the imitation rock, egg, flowers, and orange as "pretend" after a verbal prompt ("Is it a real orange [rock] or a pretend orange [rock]?" or, "Is it a pretend orange [rock] or a real orange [rock]?"), they were usually either silent when the second control question (" What is it made of?") or the Reality question was asked, repeated the apparent identity of the object, or made irrelevant responses (e.g., "It's an elevator."). In addition, for three of the subjects (Brian, Bruce, and Steve), identification of the objects as real or pretend (after the investigator failed to elicit an answer to the second control question) was dependent on the word order used by the investigator. For example, when the Orange task was administered to these subjects during the pretest, these subjects said that the orange was "pretend" when asked if it was "real or pretend." However, when the same task was re-administered as a post-test, these same subjects said that the orange was "real" after being asked if it was "pretend or real." Hence, it was apparent that Bruce, Brian and Steve consistently responded to only the last word the investigator said, with no apparent regard to the actual object.

The most parsimonious explanation for the inability of the subjects to understand the Object Identity tasks is that the language used in these tasks was considerably more difficult than that used in the Color and Size tasks. Whereas the Appearance and Reality questions in the Color and Size tasks were straightforward ("What color does it *look*?" ["Which one *looks* bigger?"], and "What color is it *really*?" ["Which one is *really* bigger?"]), the Appearance and Reality questions used in the Object Identity tasks were far more complex ("When you first saw this over here like this, what did you think it was?", and "What is it *really*?"). Contrary to the Color and Size tasks, the Object Identity tasks required the subjects to understand deitic linguistic forms such as "you" and "here", along with the juxtaposition of past and present tense markers, and the ordinal term "first." Deixis, the process of using and understanding the speaker's perspective (Cox, 1991; Owens, 1984), and seen in the comprehension and use of words like "this," "that," "here," "there," "me," "you," "near" and "far," etc., begins to develop in nonhandicapped children with MLU's of approximately 3.0 to 3.75. Although two of the subjects in the current study had MLU's exceeding 4.0, their impaired general language competencies may have interfered with their understanding of the task requirements. For Brian, Bruce, and Steve, whose MLU's were well below 3.0 (and who demonstrated more pervasive language difficulties), the format of the Object Identity tasks may have prevented them from understanding the requirements of the tasks at all.

There is also a strong possibility that the plastic orange used in the Object Identity pretest and post-test confused the subjects. When the subjects were asked the first control question ("What is this?") after the orange was revealed to the subjects, all except Steve replied "orange" on at least one presentation of the task. The absence of an indefinite article in their identification of the object (e.g., a or an), leaves open the possibility that the subjects were reporting the color of the object, instead of its identity. Still, given that the subjects were able to correctly identify the other objects' apparent identity used in the tasks, the problem presented by the use of the plastic orange was minimized.

Another aspect of the Object Identity tasks may also have contributed to the subjects' difficulty with these tasks. Whereas the appearance of the objects used in the Color and Size tasks was physically altered through the use of a filter or plexiglass, nothing was used to alter the appearance of the rock or the orange. Hence, the possibility existed that the physical change of the stimuli in the Color and Size tasks assisted the subjects in making the mental shift between the appearance of the objects and their reality. Conversely, without the physical change between the rock's or the orange's appearance and reality, the mental shift may have been much more difficult for the subjects, unattainable, in fact, for the subjects in this study.

Evaluation of Instructional Methodology

Effectiveness of Direct Instruction Procedure

The method of instruction used in the current study was Direct Instruction. The investigator believed that a structured method of example presentation would clarify the task requirements, and help the subjects in learning the appearance-reality distinction. The results of the current study suggested that for most subjects, the Direct Instructional method was adequate, since most subjects were able to reach criterion in at least one task domain. Because Brian was able to reach criterion on the Color tasks, his failure to reach criterion on the Size tasks may be attributable to an inability to understand the appearance-reality distinction concept as applied to the property of size, rather than a limitation of the instructional methodology. That is, since the instructional procedure assisted Brian in reaching criterion on the Color tasks, it may have been adequate for the Size tasks as well if he truly understood the concept. As noted in Chapter Four, Brian's performance during instruction on the Size tasks suggested that he would probably have reached criterion with additional instructional sessions.

Despite the apparent adequacy of the Direct Instruction method in teaching the appearance-reality distinction, it may not have been the best method for an extended intervention period. The repetitiveness of the task presentation resulted in rote responding by most subjects at some point during the study. While one subject used the presentation format to "teach" an imaginary friend, other subjects completed sentences when the investigator paused, or, upon seeing the equipment, spontaneously began reciting the modelling phase of instruction before they were able to demonstrate their understanding of the concept being taught. Even Steve, a subject who did not achieve criterion on either task, tended to complete the investigator's statements with no apparent understanding of the underlying appearance-reality distinction concept. Thus, the methodology allowed the subjects to learn the syntactical format of instruction easily, but obscured their understanding of the concept to a certain extent. That is, although it *appeared* that the subjects understood the concept because of their syntactic ability, it became apparent upon further questioning, that the concept was *not* actually understood.

Effectiveness of Task Presentation Method

As mentioned earlier, the Color tasks appeared to be easier than the Size tasks for Bruce and Brian. The differential difficulty between the task domains may also have been a function of the instructional method. During Color task instruction, the true color of the object concerned was altered gradually by the filter, with the true color continuously being brought to the subjects' attention as the filter changed the object's apparent color. In the initial phases, and in the initial instructional sessions, the color of the object never "became" another color suddenly. In addition, the object's apparent color was not changed in a consistent fashion. That is, sometimes the right side of the object was covered first, sometimes the left side, sometimes the top, and sometimes the bottom was covered first. This procedure may have assisted the subjects in understanding that the true color of the object had not really changed. In the brief training period included in Flavell et al.'s (1986) study, the researchers altered each object's color suddenly. The filter was either over the object or it was not. Whereas the subjects in Flavell et al.'s (1986) study did not improve in their ability to identify the true color of the objects, four of the five subjects in the current study were eventually able to make this discrimination. Thus, the current study's method of instruction used in the current study may have made the appearance-reality distinction on the Color tasks easier to comprehend.

On the other hand, the Size task procedure prohibited a gradual change of the apparent size of the objects concerned. The distortion that results when a magnifying glass is placed over an object made a gradual change of the objects' size impossible. Thus, objects used in the Size tasks were presented to the subjects in either their "true" state, or in their "apparent" state, after magnification. For Nicky, Bruce and Brian, the lack of a gradual change may have made the concept more difficult for them to understand.

Effectiveness of Reinforcement Procedure

Based on Braine and Shanks' (1965) finding that increased information feedback (in the form of winning and losing poker chips exchangeable for M&M candies contingent on correct responding) resulted in performance improvement, the investigator included a similar procedure in the current study. Hence, secondary reinforcement was introduced during instruction in the Color and Size task domains to subjects who had not reached criterion by Session 16. However, the addition of the reinforcement procedure made a difference in the responding of only one subject. Specifically, when such reinforcement was used with Nicky on the Size tasks, she completed Phase 4 of the intervention procedure immediately (Figure 4.4). The effectiveness of reinforcement on Nicky's acquisition of the Size tasks is also evident in Figure 5.2. Criterion was reached in the testing session subsequent to the introduction of reinforcement. Although the effectiveness of reinforcement is less certain for the Color tasks when one considers Nicky's performance during Phase 4 of intervention (Figure 4.2), examination of Figure 5.1 suggests that the reinforcement may have positively influenced Nicky's responding.

Contrary to Nicky's performance, secondary reinforcement made no difference in the ability of Steve to reach criterion on the Color tasks, or for Brian and Steve to reach criterion on the Size tasks. Neither of these subjects was able to complete the A-R tasks correctly or consistently in the respective task domains. The secondary reinforcement procedure was considered unnecessary for Brian, David and Bruce on the Color tasks, and for David on the Size tasks, since criterion was achieved very quickly by these subjects (within 6 intervention sessions), with verbal feedback alone.

Effectiveness of Question Format

In his appearance-reality distinction study, Baron-Cohen (1989a) did not randomize the order of the Appearance and Reality questions "because the pragmatics of the Reality question are such that it assumes the object's appearance has already been considered" (p. 592). Thus, Baron-Cohen (1989a) consistently asked the Appearance question first, and the Reality question second. The results of the current study, however, suggest that randomization of the question order is necessary. As seen in Bruce's data, the question order made a significant difference in his ability to answer correctly the Appearance and Reality questions. When the Appearance question was asked first on the Color A-R tasks, Bruce was able to complete the task correctly almost 100% of the time (apart from Session 1 there were errors on only two trials for the duration of the study). However, when the Reality question was asked first, Bruce continued to report the apparent color of the object, followed by the true color of the object when the Appearance question was subsequently asked.

In an intervention study in which feedback on performance is given, the investigator has no way of knowing if the subjects have acquired an understanding of the A-R distinction concept if the Appearance and Reality questions are always asked in a prescribed sequence. The subjects may simply be employing a strategy that requires no deeper understanding of the concept. For example, such a strategy may entail learning a rote response, one that may have been modelled by the investigator on numerous occasions, as in the present study. Validity of the results in this situation must therefore be questioned. The results of the current study demonstrated autistic children's ability to learn such rote responses. By randomization of the questions, however, rote responses were detected and controlled for, which consequently increased the study's validity.

For the subjects who ultimately reached criterion on the Color tasks, with the exception of Brian, the use of the alternate question format seemed to make a difference during baseline testing. Specifically, although Nicky and David made phenomenist errors when the standard question format was used, Nicky was able to complete 55% of the items correctly when the alternate format was used, while David was able to complete 86% of the items administered in this format. Although Bruce was able to correctly complete 100% of the tasks presented using this format, his responses were suspect for the reasons mentioned above. Brian alone, of the subjects who ultimately demonstrated understanding of the concept, was unable to complete any of the tasks correctly when the alternate format was used.

Interestingly, however, the use of the alternate format when administering the Size tasks made no significant difference in any of the subjects' ability to complete the tasks correctly. Although three subjects (Nicky, Brian and Steve) were able to complete at least one task correctly when the alternate format was used during baseline, their responses did not exceed chance. However, although unable to complete the tasks using the alternate format, Nicky, Bruce, and David did achieve criterion on the Size tasks on both the teaching tasks and the generalization tasks. It would seem, then, that the ability to complete A-R tasks using the alternate format is not a prerequisite to understanding the appearance-reality distinction, contrary to Flavell et al.'s (1986) and Baron-Cohen's (1989a) suppositions. On the contrary, the current data indicate that the alternate format is simply another way of asking the Reality question, albeit perhaps easier for some of the subjects to understand. If a correct answer to the question "If I take this away what color will it be? [which one will be bigger?]," was truly a pre-requisite to understanding appearance-reality tasks, none of the subjects should have been able to attain criterion in the Size tasks, and Brian should not have been able to attain criterion in the Color tasks.

Limitations of the Current Study and Suggestions for Future Research

The Role of Language

Although the results of this study provide some support regarding the possibility of remediating a theory of mind deficit, their interpretation cannot be divorced from the linguistic complexity of the appearance-reality tasks. The most prominent difficulty encountered in the current study was the impossibility of determining the extent to which language interfered with the demonstration of the subjects' understanding of the A-R distinction concept. The ability of the subjects to understand the terms "looks like," and "really" was only one of many linguistic difficulties encountered.

The majority of theory of mind studies conducted with autistic children (e.g., Baron-Cohen, 1989a, Baron-Cohen et al., 1986; Leslie & Frith, 1988, Prior, Dahlstrom, & Squires, 1990) have used the PPVT-R (or the British Picture Vocabulary Scale) exclusively to determine a verbal mental age. Although the central role of language in theory of mind tasks has been acknowledged (e.g., Leslie & Frith, 1988; Prior et al., 1990), it has not been adequately controlled for. For example, Leslie and Frith (1988), in comparing autistic and specifically language impaired (SLI) children's performance on false-belief tasks found that most SLI children passed these tasks while most autistic children did not pass them. However, the pragmatic language competencies between these groups of children may have been very different, and undetectable by the picture vocabulary test administered to them. Failure on the false-belief tasks may have been attributable to pragmatic language difficulties rather than to a failure employ a theory of mind.

Given the language-based nature of theory of mind tasks, more comprehensive assessment of autistic children's language competencies is vital, if these children are to be included in theory of mind research. Because these cognitive tasks can only be proposed and expressed through forms of communicative intentions, the language requirements for task presentation and comprehension must be given a more prominent role. A receptive vocabulary test cannot assess autistic children's level of language functioning accurately.

Whereas some autistic children develop substantial vocabularies (Boucher, 1989), their pragmatic language development is very heterochronous, and their language has been found to be quantitatively and qualitatively different from that of normally developing children (Wetherby & Prutting, 1984). Thus, autistic children selected as subjects for theory of mind studies must be assessed appropriately to determine if they possess the pragmatic skills necessary for completion of the tasks.

The results of the current study suggested that a critical receptive and expressive language level exists that influences the speed and accuracy of acquisition of the appearance-reality distinction tasks. In this study, the two subjects with MLU's *and* receptive vocabulary age-equivalents exceeding 4.0 (Nicky and David), reached criterion on the Color and Size tasks faster than did the other subjects in the study (refer to Table 3.1). The subject unable to reach criterion in either domain despite 10 intervention sessions (Steve) had an MLU and receptive vocabulary age-equivalent well below 4.0. The remaining subjects, although equivalent to Steve in terms of MLU, had higher receptive vocabulary age-equivalents, and attained criterion in at least one domain after a minimum of 5 intervention sessions. Although there were not enough subjects in this study to state conclusively, it appears that a minimum MLU and receptive vocabulary age-equivalent of 4.0 is the critical point at which the requirements of appearance-reality distinction tasks can be learned with relative ease. Because MLU's of 4.0 appear in normally developing children at about 4.0 years of age, concurrent with the ability to complete theory of mind tasks (Astington & Gopnik, 1991; Flavell et al., 1986; Leslie, 1987, 1988), the MLU and receptive vocabulary age-equivalent requirement is a logical one. Future studies in the area could determine the MLU's and receptive vocabularies of pre-schoolers ranging in age from 2.5 to 4.0 and assess the length of time required to reach criterion on appearance-reality tasks when instruction is provided. Similarly, studies with larger samples of autistic children could be conducted to determine the validity of the observation, and to compare results between autistic and nonautistic populations.

Relationship of Success on Appearance-Reality Tasks to Other Theory of Mind Tasks and to Social Cognition

The current study was very limited in scope, in that it only looked at acquisition of appearance-reality distinction tasks in isolation from other theory of mind tasks, and the broader social cognition implications of a theory of mind deficit. Various studies, as described in Chapter Two, have noted autistic children's deficits in the areas of false-belief attribution and understanding of causes of emotions. However, interventions directed at these deficits have not been published to date.

Extensions of the present research could investigate numerous relationships between different aspects of theory of mind. First, the relatedness between the acquisition of the appearance-reality distinction concept in a research context and its more practical application of understanding how situations and objects in children's environments can appear to be one thing but may be something else, could be addressed. It is unlikely that the subjects of the current study will be required to demonstrate their understanding that a plastic shape does not really change its color when placed behind a color filter. It is the application of the concept to social situations that is of primary interest (Flavell et al., 1986). Thus, an extension of this study using more ecologically valid situations (e.g., the understanding of misexpectations and misperceptions as they relate to social interaction) would be of benefit.

Secondly, future research could compare autistic children's acquisition of understanding of appearance-reality tasks to their acquisition of understanding of false-belief and causes of emotions tasks. Several children in the current study reached criterion in one or both task domains on the appearance-reality tasks. However, neither this study, nor any other study (to this investigator's knowledge) has compared autistic children's performance on different aspects of theory of mind (e.g., false-belief, causes of emotion). Research comparing performance in these different areas may reveal some interesting relationships between the acquisition of these concepts in the autistic population.

Finally, studies in which investigators compare normal and autistic children's social cognitive abilities as observed and assessed in naturalistic situations with their ability to complete theory of mind tasks could be conducted. Such research may help in clarifying the relationship between the domains, and the role played by deficits in social cognition in autistic children's understanding of theory of mind concepts.

Symbolic Play and Theory of Mind

In Chapter Two the role of symbolic play as a precurser in the development of a theory of mind was discussed. Leslie (1987) proposed that children incapable of pretense, will also be impaired in their ability to form the metarepresentations required for a fully functional theory of mind. Although autistic children's capacity for symbolic play has been disputed (e.g., Lewis & Boucher, 1988), there appears to be agreement that play demonstrated by autistic children is qualitatively different than that of normally developing children of the same age (Baron-Cohen, 1987; Riguet et al., 1981;Wing et al., 1977). However, the connection between symbolic play and theory of mind has not been systematically explored.

Evidence of symbolic (and functional) play was observed in most of the subjects participating in the current study when the language sample was obtained. In the collection of the language sample, the subjects were allowed to play as they wished with miniature kitchen toys and plastic food. All of the children who eventually attained criterion in at least one domain demonstrated symbolic play. Most of the children poured imaginary substances from one pot to another, made noises representing food sizzling on the stove, mixed imaginary food in a bowl, or pretended to burn themselves by touching a "hot" element. Steve alone played with the toys in a stereotypic manner. When presented with the kitchen toys he spun the top to a pot and spun a plastic tomato. At no time was any functional or symbolic play observed.

Although not systematically investigated in the current study, the noted observations provided limited evidence that symbolic play may be a precurser to success on theory of mind tasks. Further investigation of symbolic play and theory of mind that takes pragmatic and semantic language ability into account, may clarify and expand on the hypothesized relationship between the two.

Summary and Conclusions

The current study was a preliminary one illustrating the potential of Direct Instruction in teaching the appearance–reality distinction to autistic children, at least as it pertains to the properties of color and size. Of the five subjects included in the study, four reached criterion on the Color tasks, with three subjects reaching criterion on the Size tasks as well. One subject failed to reach criterion in either task domain. There appeared to be a relationship between the task domains, in that all subjects who reached criterion on the Size tasks, also reached criterion on the Color tasks, with criterion on the Color tasks being achieved first. The two exceptions to this trend were David and Nicky. Whereas Nicky achieved criterion in the two domains within a day of each other, David achieved criterion in both task domains simultaneously. With respect to the Object Identity generalization tasks, none of the subjects were able to answer the second control question on the Object Identity tasks correctly. Since the children did not seem to understand the requirements of the Object Identity tasks, their performance on these tasks could not be interpreted.

Three hypotheses could account for criterion in the Color and Size task domains being reached by the subjects concerned.

1. Prior to instruction, the subjects did not understand the appearance-reality distinction concept, and did actually learn the concept through the instruction provided.
2. The subjects did, in fact, have an intuitive understanding of the concept prior to instruction, but because of general language comprehension difficulties, did not fully understand the requirements of the present tasks until instruction had taken place.

3. The subjects possessed an intuitive understanding of the appearance-reality distinction concept, but did not understand the lexical terms "looks like" and "really", the meaning of which was learned in the course of instruction.

If the second or third hypotheses accounted for the behavior change of the subjects, and the subjects simply did not understand the requirements of the task or the terms used, it would seem logical that any subject who achieved criterion in one of the task domains would achieve criterion in the other domain as well, once the task requirements and the terms were understood, particularly if the A-R distinction appears across domains simultaneously as Flavell et al. (1986) suggest. However, one of the four subjects attaining criterion on the Color tasks (Brian) did not attain criterion on the Size tasks, and none of the subjects completed the Object Identity tasks correctly. The subjects' failure on the Object Identity tasks, however, may have been due to the other reasons previously discussed.

If, on the other hand, the first hypothesis accounted for the subjects' change in behavior, and understanding of the appearance-reality distinction concept was attained in the course of the study, concept understanding for some subjects may not have been transferable to tasks involving different object properties. Although the Color and Size tasks were similar enough in their requirements and attributes to enable some subjects to attain criterion in both task domains, the differences between the two types of tasks may have been sufficient to inhibit generalization of the concept to the Size tasks for Brian, and generalization to the Object Identity tasks for all of the subjects. It should be noted that all subjects who attained criterion in a task domain were able to complete the generalization tasks correctly within that domain as well. Thus, application of the A-R concept within a domain was not limited to particular objects. Once concept understanding was demonstrated, it remained stable regardless of the object used and was generalizable to tasks within the domain that were not taught directly.

Baron-Cohen (1988) has postulated that the observed theory of mind deficit in autistic children supports a cognitive interpretation of the primary deficit in autism. Hobson (1986a, 1986b, 1990) on the other hand, has stated that a primary affective deficit underlies the autistic syndrome and autistic children's failure to understand theory of mind tasks. As Wetherby and Prutting (1984) has suggested, however, the developmental interplay of the linguistic, cognitive, and social domains may make it impossible to identify the primary disorder in autism. Throughout the current study, the subjects exhibited receptive and expressive language comprehension problems which have led this investigator to agree with Wetherby and Prutting (1984), in that it is virtually impossible to extricate the cognitive aspects of the appearance-reality distinction tasks from the linguistic aspects of the tasks.

Research on theory of mind has shown that the acquisition of conceptual perspective-taking is an important aspect of cognitive development. Similarly, apparent deficits in the acquisition of such a theory in autistic children is an important area of investigation. However, for research on autistic children's theory of mind deficits to be valid, any deficits must be considered within the total context of the autistic child and his/her difficulties with pragmatic language and communication, social relatedness, and abstract and symbolic thought. Once the total context is appreciated, effective interventions can be implemented.

REFERENCES

American Psychiatric Association. (1980). *Diagnostic and statistical manual of mental disorders,* (3rd ed.,). Washington, DC: Author.

American Psychiatric Association. (1987). *Diagnostic and statistical manual of mental disorders,* (3rd ed., rev.). Washington, DC: Author.

Astington, J. W., & Gopnik, A. (1988). Knowing you've changed your mind: Children's understanding of representational change. In J. W. Astington, P. L. Harris, & D. R. Olson (Eds.), *Developing theories of mind* (pp. 193-206). Cambridge: Cambridge University Press.

Astington, J. W., & Gopnik, A. (1991). Theoretical explanations of children's understanding of the mind. *British Journal of Developmental Psychology, 9,* 7-31.

Avis, J., & Harris, P. L. (1991). Belief-desire reasoning among Baka children: Evidence for a universal conception of mind. *Child Development, 62,* 460-467.

Barlow, D. H., & Hersen, M. (1984). *Singe case experimental designs: Strategies for studying behavior change* (2nd ed.). New York: Pergamon Press.

Baron, M. G., Groden, G., & Cautela, J. R. (1988). Behavioral programming: Expanding our clinical repertoire. In G. Groden & M. G. Baron (Eds.), *Autism: Strategies for change* (pp. 49-73). New York: Gardner Press.

Baron-Cohen, S. (1987). Autism and symbolic play. *British Journal of Developmental Psychology, 5,* 139-148.

Baron-Cohen, S. (1988). Social and pragmatic deficits in autism: Cognitive or affective? *Journal of Autism and Developmental Disorders, 30,* 931-933.

Baron-Cohen, S. (1989a). Are autistic children "behaviorists"? An examination of their mental-physical and appearance-reality distinctions. *Journal of Autism and Developmental Disorders, 19,* 579-600.

Baron-Cohen, S. (1989b). Joint-attention deficits in autism: Towards a cognitive analysis. *Development and Psychopathology, 1,* 185-189.

Baron-Cohen, S. (1989c). Perceptual role taking and protodeclarative pointing in autism. *British Journal of Developmental Psychology, 7,* 113-127.

Baron-Cohen, S. (1989d). The autistic child's theory of mind: A case of specific developmental delay. *Journal of Psychology and Psychiatry, 30,* 285-297.

Baron-Cohen, S. (1989e). The theory of mind hypothesis of autism: A reply to Boucher. *British Journal of Disorders of Communication, 24,* 199-200.

Baron-Cohen, S. (1990a). Autism: A specific cognitive disorder of 'mind-blindness'. *International Review of Psychiatry, 2,* 81-90.

Baron-Cohen, S. (1990b). Instructed and elicited play in autism: A reply to Lewis & Boucher. *British Journal of Developmental Psychology, 8,* 207.

Baron-Cohen, S. (1991a). Do people with autism understand what causes emotion? *Child Development, 62,* 385-395.

Baron-Cohen, S. (1991b). The theory of mind deficit in autism: How specific is it? *British Journal of Developmental Psychology, 9,* 301-314.

Baron-Cohen, S., Leslie, A. M. & Frith, U. (1985). Does the autistic child have a 'theory of mind'? *Cognition, 21,* 37-46.

Baron-Cohen, S., Leslie, A. M. & Frith, U. (1986). Mechanical, behavioural and Intentional understanding of picture stories in autistic children. *British Journal of Developmental Psychology, 4,* 113-125.

Bartak, L., & Pickering, G. (1976). Aims and methods of teaching. In M. P. Everard (Ed.), *An approach to teaching autistic children* (pp. 79-98). Oxford: Pergamon Press.

Becker, W. C., & Carnine, D. W. (1980). Direct Instruction: An effective approach to educational intervention with the disadvantaged and low performers. In B. B. Lahey, & A. E. Kazdin (Eds.), *Advances in clinical child psychology, Vol. 3* (pp. 429-473). New York: Plenum Press.

Bettelheim, B. (1959). Joey: A "mechanical boy." *Scientific American, 200* (3), 116-127.

Bettelheim, B. (1967). *The empty fortress*. New York: Free Press.

Boucher, J. (1989). The theory of mind hypothesis of autism: Explanation, evidence and assessment. *British Journal of Disorders of Communication, 24,* 181-198.

Boucher, J., & Lewis, V. (1990). Guessing or creating? A reply to Baron-Cohen. *British Journal of Developmental Psychology, 8,* 205-206.

Braine, M. D. S., & Shanks, B. L. (1965). The development of conservation of size. *Journal of Verbal Learning and Verbal Behavior, 4,* 227-242.

Bretherton, I., & Beeghly, M. (1982). Talking about internal states: The acquisition of an explicit theory of mind. *Developmental Psychology, 18,* 906-921.

Brown, R. (1973). *A first language: The early stages*. Cambridge MA: Harvard University Press.

Bryson, S. E., Clark, B. S., & Smith, I. M. (1988). First report of a Canadian epidemiological study of autistic syndromes. *Journal of Child Psychology and Psychiatry, 29,* 433-445.

Carnine, D. W., & Becker, W. C. (1982). Theory of instruction: Generalisation issues. *Educational Psychology, 2,* 249-262.

Case, R. (1985). A developmentally based approach to the problem of instructional design. In S. F. Chipman, J. W. Segal, & R. Glaser (Eds.). *Thinking and learning skills: Vol.*

2. *Research and open questions* (pp. 545-562). Hillsdale, NJ: Lawrence Erlbaum Associates.

Cohen, D. (1983). *Piaget: Critique and reassessment.* London: Croom Helm.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20,* 37-46.

Courchesne, E. (1989). Neuroanatomical systems involved in infantile autism. In G. Dawson (Ed.), *Autism: Nature, diagnosis, and treatment* (pp. 119-173). New York: The Guilford Press.

Courchesne, E. (1991a, July). *A new model of brain and behavior development in infantile autism.* Paper presented at the conference of the Autism Society of America, Indianapolis, IN.

Courchesne, E. (1991b). Neuroanatomic imaging in autism. *Pediatrics, 87* (5), (Supplement) 781-790.

Cox, M. V. (1991). *The child's point of view* (2nd ed.).London: Harvester Wheatsheaf.

de Gelder, B. (1987). On not having a theory of mind. *Cognition, 27,* 285-290.

Dunlap, G. (1984). The influence of task variation and maintenance tasks on the learning and affect of autistic children. *Journal of Experimental Child Psychology, 37,* 41-64.

Dunlap, G., & Koegel, R. L. (1980). Motivating autistic children through stimulus variation. *Journal of Applied Behavior Analysis, 13,* 619-627.

Dunlap, G., Koegel, R. L., & Egel, A. L. (1979). Autistic children in school. *Exceptional Children, 45,* 552-558.

Dunn, L. & Dunn, L. (1981). *Peabody Picture Vocabulary Test-Revised.* Circle Pines, MN: American Guidance Service.

Eisenmajer, R., & Prior, M. (1991). Cognitive linguistic correlates of 'theory of mind' ability in autistic children. *British Journal of Developmental Psychology, 9,* 351-364.

Engelmann, S., & Carnine, D. (1982). *Theory of instruction: Principles and application.* New York: Irvington Publishers.

Englert, C. S. (1984). Effective direct instruction practices in special education settings. *Remedial and Special Education, 5*(2), 38-47.

Fein, D., Pennington, B., Markowitz, P., Braverman, M., & Waterhouse, L. (1986). Toward a neuropsychological model of infantile autism: Are the social deficits primary?. *Journal of the American Academy of Child Psychiatry, 25,* 198-212.

Flavell, J. H. (1988). The development of children's knowledge about the mind: From cognitive connections to mental representations. In J. W. Astington, P. L. Harris, & D. R. Olson (Eds.), *Developing theories of mind* (pp. 244-267). Cambridge: Cambridge University Press.

Flavell, J. H., Everett, B. A., Croft, K., & Flavell, E. R. (1981). Young children's knowledge about visual perception: Further evidence for the Level 1-Level 2 distinction. *Developmental Psychology, 17,* 99-103.

Flavell, J. H., Flavell, E. R., & Green, F. L. (1983). Development of the appearance-reality distinction. *Cognitive Psychology, 15,* 95-120.

Flavell, J. H., Flavell, E. R., & Green, F. L. (1987). Young children's knowledge about the apparent-real and pretend-real distinctions. *Developmental Psychology, 23,* 816-822.

Flavell, J. H., Green, F. L., & Flavell, E. R. (1986). Development of knowledge about the appearance-reality distinction. *Monographs of the Society for Research in Child Development, 51* (1, Serial No. 212)

Flavell, J. H., Zhang, X-D, Zou, H., Dong, Q., & Qi, S. (1983). A comparison between the development of the appearance-reality distinction in the People's Republic of China and the United States. *Cognitive Psychology, 15,* 459-466.

Frith, U. (1989a). A new look at language and communication in autism. *British Journal of Disorders of Communication, 24,* 123-150.

Frith, U. (1989b). *Autism: Explaining the enigma.* Oxford: Basil Blackwell.

Gallagher, J. J., & Wiegerink, R. (1976). Educational strategies for the autistic child. *Journal of Autism and Childhood Schizophrenia, 6,* 15-26.

Gerston, R. (1985). Direct instruction with special education students: A review of evaluation research. *Journal of Special Education, 19,* 41-58.

Gillberg, C. (1990). Autism and pervasive developmental disorders. *Journal of Child Psychology and Psychiatry, 31,* 99-119.

Ginsburg, H., & Opper, S. (1979). *Piaget's theory of intellectual development* (2nd ed.). Englewood Cliffs, NJ: Prentice-Hall.

Gopnik, A., & Astington, J. W. (1988). Children's understanding of representational change and its relation to the understanding of false belief and the appearance-reality distinction. *Child Development, 59,* 26-37.

Hammes, J. & Langdell, T. (1981). Precursors of symbol formation and childhood autism. *Journal of Autism and Developmental Disorders, 11,* 331-346.

Hartmann, D. P. (1977). Considerations in the choice of interobserver reliability estimates. *Journal of Applied Behavior Analysis, 10,* 103-116.

Hermelin, B. & O'Conner, N. (1970). *Psychological experiments with autistic children.* Oxford: Pergamon Press.

Hobson, R. P. (1984). Early childhood autism and the question of egocentrism. *Journal of Autism and Developmental Disorders, 14,* 85-104.

Hobson, R. P. (1986a). The autistic child's appraisal of expressions of emotion. *Journal of Child Psychology and Psychiatry, 27,* 321-342.

Hobson, R. P. (1986b). The autistic child's appraisal of expressions of emotion: A further study. *Journal of Child Psychology and Psychiatry, 27,* 671-680.

Hobson, R. P. (1989). Emotion-related and abstract concepts in autistic people: Evidence from the British Picture Vocabulary Scale. *Journal of Autism and Developmental Disorders, 19,* 601-623.

Hobson, R. P. (1990). On acquiring knowledge about people and the capacity to pretend: Response to Leslie (1987). *Psychological Review, 97,* 114-121.

Hobson, R. P. (1991). Against the theory of 'theory of mind'. *British Journal of Developmental Psychology, 9,* 33-51.

Hogrefe, F. J., Wimmer, H., & Perner, J. (1986). Ignorance versus false belief: A developmental lag in attribution of epistemic states. *Child Development, 57,* 567-582.

Kameenui, E. J., & Simmons, D. C. (1990). *Designing instructional strategies: The prevention of academic learning problems.* Columbus, OH: Merrill Publishing Company.

Kanner, L. (1985). Autistic disturbances of affective contact. In A. M. Donnellan (Ed.), *Classic readings in autism* (pp. 11-50). New York: Teachers College Press.

Kazdin, A. E. (1981). Drawing valid inferences from case studies. *Journal of Consulting and Clinical Psychology, 49,* 183-192.

Kazdin, A. E. (1982). *Single-case research designs: Methods for clinical and applied settings.* New York: Oxford University Press.

Kistner, J., & Robbins, F. (1986). Brief report: Characteristics of methods of subject selection and description in research on autism. *Journal of Autism and Developmental Disorders, 16,* 77-82.

Koegel, R. L., & Egel, A. L. (1979). Motivating autistic children. *Journal of Abnormal Psychology, 88,* 418-426.

Koegel, R. L., Dunlap, G., & Dyer, K. (1980). Intertrial interval duration and learning in autistic children. *Journal of Applied Behavior Analysis, 13,* 91-99.

Leiter, R. G. (1979). *Leiter International Performance Scale: Instruction manual.* Chicago: Stoelting.

Leslie, A. M. (1987). Pretense and representation: The origins of "theory of mind". *Psychological Review, 94,* 412-426.

Leslie, A. M. (1988). Some implications of pretense for mechanisms underlying the child's theory of mind. In J. W. Astington, P. L. Harris, & D. R. Olson (Eds.), *Developing theories of mind* (pp. 19-46). Cambridge: Cambridge University Press.

Leslie, A. M., & Frith, U. (1988). Autistic children's understanding of seeing, knowing and believing. *British Journal of Developmental Psychology, 6,* 315-324.

Leslie, A. M., & Frith, U. (1990). Prospects for a cognitive neuropsychology of autism: Hobson's choice. *Psychological Review, 97,* 122-131.

Levine, M. N. (1982). *Leiter International Performance Scale: A handbook.* Los Angeles: Western Psychological Services.

Lewis, V., & Boucher, J. (1988). Spontaneous, instructed and elicited play in relatively able autistic children. *British Journal of Developmental Psychology, 6,* 325-339.

Lotter, V. (1985). Epidemiology of autistic conditions in young children: Part 1. Prevalence. In In A. M. Donnellan (Ed.), *Classic readings in autism* (pp. 105-134). New York: Teachers College Press.

Lovaas, O. I. (1977). *The autistic child: Language development through behavior modification.* New York: Irvington Publishers.

Lovaas, O. I. (1981). *Teaching developmentally disabled children: The me book.* Austin, TX: Pro-ed.

Maltz, A. (1981). Comparison of cognitive deficits among autistic and retarded children on the Arthur Adaptation of the Leiter International Performance Scales. *Journal of Autism and Developmental Disorders, 11,* 413-426.

Marvin, R. S., Greenberg, M. T., & Mossler, D. G. (1976). The early development of conceptual perspective taking: Distinguishing among multiple perspectives. *Child Development, 47,* 511-514.

Matey, C. (1985). Leiter International Performance Scale. In D. J. Keyser, & R. C. Sweetland (Eds.), *Test Critiques, Vol. 1* (pp. 411-420). Kansas: Test Corporation of America.

Moore, J. (1986). Direct Instruction: A model of instructional design. *Educational Psychology, 6,* 201-229.

Mundy, P., Sigman, M., Ungerer, J., & Sherman, T. (1987). Nonverbal communication and play correlates of language development in autistic children. *Journal of Autism and Developmental Disorders, 17,* 349-364.

O'Neill, R. E., & Dunlap, G. (1989). DI principles work with autistic children. *Direct Instruction News, 8* (2), 29-31.

Olson, D. (1988). On the origins of beliefs and other intentional states in children. In J. W. Astington, P. L. Harris, & D. R. Olson (Eds.), *Developing theories of mind* (pp. 414-426). Cambridge: Cambridge University Press.

Olson, D. R., Astington, J. W., & Harris, P. L. (1988). Introduction. In J. W. Astington, P. L. Harris, & D. R. Olson (Eds.), *Developing theories of mind* (pp. 1-15). Cambridge: Cambridge University Press.

Owens, R. E. (1984). *Language development: An introduction.* Columbus OH: Charles E. Merrill Publishing Company.

Perner, J. (1988). Higher-order beliefs and intentions in children's understanding of social interaction. In J. W. Astington, P. L. Harris, & D. R. Olson (Eds.), *Developing theories of mind* (pp. 271-294). Cambridge: Cambridge University Press.

Perner, J., & Wimmer, H. (1985). "John thinks that Mary thinks that..." Attribution of second-order beliefs by 5- to 10-year-old children. *Journal of Experimental Child Psychology, 39,* 437-471.

Perner, J., Frith, U., Leslie, A. M., & Leekam, S. R. (1989). Exploration of the autistic child's theory of mind: Knowledge, belief, and communication. *Child Development, 60,* 689-700.

Perner, J., Leekam, S. R., & Wimmer, H. (1987). Three-year-olds' difficulty with false belief: The case for a conceptual deficit. *British Journal of Developmental Psychology, 5,* 125-137.

Piaget, J. (1962). *Play, dreams, and imitation in childhood.* (C. Gattegno & F. M. Hodgson, Trans.). London: Routledge & Kegan Paul Ltd. (Original work published 1957).

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *The Behavioral and Brain Sciences, 4,* 515-526.

Prior, M., Dahlstrom, B., & Squires, T. (1990). Autistic children's knowledge of thinking and feeling states in other people. *Journal of Child Psychology and Psychiatry, 31,* 587-601.

Riguet, C. B., Taylor, N. D., Benaroya, S., & Klein, L. S. (1981). Symbolic play in autistic, Down's and normal children of equivalent mental age. *Journal of Autism and Developmental Disorders, 11,* 439-448.

Rimland, B. (1985). The etiology of infantile autism: The problem of biological versus psychological causation. In A. M. Donnellan (Ed.), *Classic readings in autism* (pp. 84-103). New York: Teachers College Press.

Rincover, A. (1978). Variables affecting stimulus fading and discriminative responding in psychotic children. *Journal of Abnormal Psychology, 87,* 541-553.

Rincover, A., & Koegel, R. L. (1975). Setting generality and stimulus control in autistic children. *Journal of Applied Behavior Analysis, 8,* 235-246.

Rincover, A., Cook, R., Peoples, A., & Packard, D. (1979). Using sensory extinction and sensory reinforcement principles for programming multiple treatment gains in autistic children. *Journal of Applied Behavior Analysis, 12*, .221-233.

Rincover, A., Newsom, C. D., Lovaas, O. I., & Koegel, R. L. (1977). Some motivational properties of sensory stimulation in psychotic children. *Journal of Experimental Child Psychology, 24*, 312-323.

Ruffman, T. K., & Olson, D. R. (1989). Children's ascriptions of knowledge to others. *Developmental Psychology, 25*, 601-606.

Russell, J., Mauthner, N., Sharpe, S., & Tidswell, T. (1991). The 'windows task' as a measure of strategic deception in preschoolers and autistic subjects. *British Journal of Developmental Psychology, 9*, 331-349.

Rutter, M. (1978). Diagnosis and definition of childhood autism. *Journal of Autism and Childhood Schizophrenia, 8*, 139-161.

Rutter, M. (1983). Cognitive deficits in the pathogenesis of autism. *Journal of Child Psychology and Psychiatry, 24*, 513-531.

Rutter, M. (1985). Concepts of autism: A review of research. In A. M. Donnellan (Ed.), *Classic readings in autism* (pp. 179-208). New York: Teachers College Press.

Rutter, M. (1987). The role of cognition in child development and disorder. *British Journal of Medical Psychology, 60*, 1-16.

Sattler, J. M. (1982). *Assessment of children's intelligence and special abilities* (2nd ed.). Boston: Allyn and Bacon.

Schopler, E., Reichler, R. J., & Lansing, M. (1980). *Individualized assessment and treatment for autistic and developmentally disabled children: Vol. 2. Teaching strategies for parents and professionals*. Baltimore: University Park Press.

Shah, A., & Holmes, N. (1985). Brief report: The use of the Leiter International Performance Scale with autistic children. *Journal of Autism and Developmental Disorders, 15*, 195-203.

Sigman, M., & Mundy, P. (1987). Symbolic processes in young autistic children. *New Directions for Child Development, 36*, 31-46.

Sigman, M., & Ungerer, J. A. (1984). Cognitive and language skills in autistic, mentally retarded, and normal children. *Developmental Psychology, 20*, 293-302.

Sigman, M., Ungerer, J. A., Mundy, P., & Sherman (1987). Cognition in autistic children. In D. J. Cohen, A. M. Donnellan, & Paul, R. (Eds.), *Handbook of autism and Pervasive Developmental Disorders*. New York: John Wiley & Sons.

Smith, M. D. (1978). *Educational psychology and its classroom applications* (2nd ed.). Boston: Allyn & Bacon.

Smith, M. L., & Glass, G. V. (1987). *Research and evaluation in education and the social sciences*. Englewood Cliffs, NJ: Prentice-Hall.

Stokes, T. F., & Baer, D. M. (1977). An implicit technology of generalization. *Journal of Applied Behavior Analysis, 10*, 349-367.

Tawney, J. W., & Gast, D. L. (1984). *Single subject research in special education*. Columbus, OH: Charles E. Merrill Publishing Company.

Tinbergen, N. & Tinbergen, E. A. (1983). *Autistic children: New hope for a cure*. Boston: George Allen & Unwin.

Wellman, H. M. (1985). The child's theory of mind: The development of conceptions of cognition. In S. Yussen (Ed.), *The growth of reflection in children* (pp. 169-206). Orlando, FL: Academic Press, Inc.

Wetherby, A. M., & Prutting, C. A. (1984). Profiles of communicative and cognitive-social abilities in autistic children. *Journal of Speech and Hearing Research, 27*, 364-377.

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition, 13*, 103-128.

Wing, L. (1980). *Autistic children* (2nd ed.). London: Constable & Co. Ltd.

Wing, L., Gould, J., Yeates, S. R., & Brierley, L.M. (1977). Symbolic play in severely mentally retarded and in autistic children. *Journal of Child Psychology and Psychiatry, 18*, 167-178.

Wulff, S. B. (1985). The symbolic and object play of children with autism: A review. *Journal of Autism and Developmental Disorders, 15*, 139-148.

APPENDIX A

THEORY OF MIND STUDY SUMMARIES

Theory of mind study summaries.

Type	*Author (Date)*	*Subject Description*	*Tasks*	*Results*	*Conclusions*
False Belief	Baron-Cohen et al. (1985)	Autistic-N=20 -mean CA=11:11 -mean nonverbal MA=9:3 -mean verbal MA=5:5 Down's-N=14 -mean CA=10:11 -mean nonverbal MA=5:11 -mean verbal MA=2:11 Normal-N=27 -mean CA=4:5	-2 doll protagonists (Sally and Anne) -Sally places marble in basket and leaves scene -Anne transfers marble to box -Sally returns -subject asked control questions and critical belief question ("Where will Sally look for the marble?") -2 trials given	-all subjects passed control questions -85% normal and 86% Down's subjects passed belief question on both trials -80% autistic failed belief question on both trials -4 autistic passed (CA 10:11-15:10, nonverbal MA 8:10-10:8, verbal MA 2:9-7:0)	-autistic don't appreciate difference between their own and doll's knowledge -as a group autistic fail to employ a theory of mind (inability to represent mental states) -constitutes an autism specific deficit
False Belief	Baron-Cohen et al. (1986)	Autistic-N=21 -mean CA=12:4 -mean nonverbal MA=9:6 -mean verbal MA=5:7 Down's-N=15 -mean CA=10:5 -mean nonverbal MA=5:9 -mean verbal MA=2:9 Normal-N=27 -mean CA=4:5	-assessed subjects' ability to arrange pictures into predetermined sequence -understanding checked in an analysis of narrations of sequences -5 types of stories -**Mechanical 1**: Objects interacting causally with each other -**Mechanical 2**: People and objects acting causally on each other -**Behavioral 1**: Single person acting in routines not requiring attribution of mental states -**Behavioral 2**: People acting in social routines, involving more than one person not requiring attribution of mental states -**Intentional**: People acting in activities requiring attribution of mental states -3 stories for each condition	-performance on mechanical and behavioral subconditions highly similar between groups (results collapsed) -autistic significantly better than Down's and normal on mechanical -autistic and normal performed identically on behavioral condition while Down's significantly worse -autistic performance significantly worse than Down's and normal on Intentional condition -Down's worse in comparison with autistic on all conditions except Intentional one	-confirmed findings that autistic show specific deficit in employing a theory of mind -no support for general "social" deficit explanation of autism -found specific deficit in social understanding -likely that at least 2 autistic subjects able to employ theory of mind -suggest that critical feature of social impairment in autism may involve specific dysfunction in conceiving of mental states

Type	*Author (Date)*	*Subject Description*	*Tasks*	*Results*	*Conclusions*
Mental States	Leslie & Frith (1988)	Autistic-N=18 -mean CA=13:10 -mean verbal MA=7:2 Specific Language Impaired -N=12 -mean CA=8:8 -mean verbal MA=6:9	**Line of Sight** -subject asked if doll could see counter placed on either side of a board **Memory for Position** -subject required to replace a marble to its original location **Limited Knowledge** -subject required to determine where a currently absent experimenter would look for a counter moved and hidden in a location different from its location when experimenter was present (prediction question) -child asked if experimenter knows current location of counter (knowing question) **False Belief** -As in Baron-Cohen (1985) but using real people including knowing question ("Does experimenter know the coin is in here?"), prediction question ("Where will experimenter look for the coin?"), and think question ("Where does experimenter think the coin is?")	**Line of Sight** -all autistic passed line of sight test **Memory for Position** -all autistic passed **Limited Knowledge** -knowing question passed by 11 and failed by 7 autistic -prediction question passed by 9 and failed by 9 autistic -to pass-must answer both correctly -therefore-8 passed and 10 failed **False Belief** -prediction question passed by 5 and failed by 13 autistic -passing or failing prediction question consistent with passing or failing prediction question -all SLI passed prediction and knowing questions -1 SLI failed thinking question -significant difference between groups -high contingency between limited knowledge and false belief tasks -no relationship between failing/passing and MA -trend for older children to perform better than younger children	-as a group autistic have difficulty with idea of mental states -language problems not cause of impairment of mental state concept -presumed that line of sight passed by employing geometric concept regarding causal connection between line of sight and target -since older children were ones who did show evidence of meta-representational understanding it is suggested that with practice or maturation autistic child can increase whatever facility he or she has with mental state concepts -autistic are specifically impaired in meta-representational capacity in a way which produces gross delay in development of theory of mind

Type	*Author (Date)*	*Subject Description*	*Tasks*	*Results*	*Conclusions*
Knowledge, Belief and Communication	Perner et al. (1989)	Autistic-N=26 -mean CA=13:6 -mean verbal MA=6:2 Specific Language Impaired --N=12 -mean CA=8:8 mean verbal MA=6:9	**Communication Test-Boxes** -objects hidden in 2 boxes after one experimenter leaves room -partial and total ignorance conditions -returned experimenter asks subject what is in each box in each condition **False-Belief Test** -smarties box with pencil -subject asked what another subject will think is in smarties box **Communication Test-Bee** -subject asked to tell previously absent experiementer what a toy bee can do upon experimenter's return -partial and total ignorance conditions **Knowledge-formation** -object hidden in cup -Other Ignorant Condition-subject saw object _Subject Ignorant Condition-experimenter saw object	-4 of 23 autistic answered false-belief prediction question correctly -SLI had little difficulty with false-belief (all but 1 passed) -all 4 autistic passing belief task and 4 others passed -other autistic failed -only 12.5% of autistic reliably took into account listener's knowledge on communication task	-supports claim that autistic are severely impaired in theory of mind -impairment not result of general mental retardation -confirms and extends previous findings -certain aspects of theory of mind more easily developed than others

Type	*Author (Date)*	*Subject Description*	*Tasks*	*Results*	*Conclusions*
2nd order false-belief	Baron-Cohen (1989d)	Autistic-N=10 -mean CA=15:3 -mean nonverbal MA=10:7 -mean verbal MA -expressive=12:2 -receptive=7:8 Down's-N=10 -mean CA=14:3 -mean nonverbal MA=6:8 -mean verbal MA -expressive=7:5 -receptive=4:7 Normal-N=10 -mean CA=7:3	-subjects shown toy village that included 4 "playpeople" -subjects told story while characters were moved around the village -asked belief question ("Where does Mary think John has gone to buy an ice-cream?"), justification question ("Why?"), reality question ("Where did John really go to buy his ice-cream?"), and memory question ("Where was the ice-cream man in the beginning?")	**Control Questions** -all subjects except 1 autistic passed all control questions **Belief Question** -90% of normal, 60% of Down's, 0% of autistic passed 2nd order belief question -Down's and normal did not differ significantly and autistic significantly worse than Down's **Justification Question** -responses coded into 1 of 3 categories of belief attribution according to whether subject took account of (a) John's *and* Mary's beliefs (2nd order) (b) John's *or* Mary's beliefs (1st order) (c) Neither of their beliefs (0 order) **Responses** **Autistic** -0-2nd order responses -4-1st order responses -5-0 order responses -1 autistic NR **Down's** -6-2nd order responses -2-1st order responses -1-0 order response -1-NR **Normal** -9-2nd order responses -1-1st order response	-autistics able to attribute beliefs at simplest level but unable to attribute beliefs at higher level -support prediction that autistic who have developed theory of mind at lower level are specifically delayed in acquisition of more complex theory of mind -minimum CA may be necessary -autistic may be delayed in development of theory of mind by 7 chronological years

Type	*Author (Date)*	*Subject Description*	*Tasks*	*Results*	*Conclusions*
Appearance-Reality Distinction	Baron-Cohen (1989a)	Autistic-N=17 -mean CA=13:8 -mean nonverbal MA=8:5 -mean verbal MA=6:9 Mentally Handicapped-N=16 -mean CA=15:4 -mean nonverbal MA=6:0 -mean verbal MA=6:5 Normal-N=19 -mean CA=5:3	**Experiment 1** -testing understanding of mental phenomena (thoughts, dreams, pretense and memories) -story with mental/physical contrast between dolls -asked judgement questions **Experiment 2** -tested concept of brain -location and function questions asked about brain and heart **Experiment 3** -appearance-reality distinction -4 A-R tasks in which object's color, size, material and identity were manipulated separately	**Experiment 1** -78.9% of normals , 68.8% mentally handicapped and 23.5% autistic passed judgement questions -autistic significantly lower than other 2 groups **Experiment 2** 84.2% normals, 68.8% mentally handicapped, and 23.5% autistic referred to mental phenomenon as one function of the brain -70.6% autistic referred to brain's role in behavior -significant difference -function question on heart passed by 82.4% autistics, 87.5% of mentally handicapped and 94.7% of normals -difference not significant **Experiment 3** -81.3% normals, 78.9% mentally handicapped and 35.3% autistic passed tasks -highly significant difference -tasks not significantly different in difficulty -autistic tended to make phenomenist errors	-autistic treat mental phenomena no differently than physical ones -autistic tended to associate brain with behavioral function which is more complex for normals -more complex answers may be due to higher MA -A-R results suggest when perceptual information contradicts own knowledge about world, autistic unable to separate these and perceptual information overrides other representations of object -autistic inability to represent object in 2 different ways simultaneously -understanding of all internal states not above autistics capabilities

Type	*Author (Date)*	*Subject Description*	*Tasks*	*Results*	*Conclusions*
Causes of Emotion	Baron-Cohen (1990)	Autistic-N=17 -mean CA=13:8 -mean nonverbal MA=8:5 -mean verbal MA=6:9 Mentally Handicapped-N=16 -mean CA=15:4 -mean nonverbal MA=6:0 -mean verbal MA=6:5 Normal-N=19 -mean CA=5:3	-subjects told story interspersing questions testing understanding of situations as cause of emotion (Situation Test) questions testing understanding of desires as cause of emotion (1st Desire Test), questions testing understanding of beliefs as cause of emotion (Belief Test), Second Desire Test to control for order effects -each test question followed by Justification Question	-autistic significantly worse than other 2 groups on Belief Test (17.6% autistic, 56.3% MH, 73.7% normals passed) -on Desire Test 57.4% autistic, 59.4% MH and 92.1% normals passed	-Desire Tests easier than Belief Test for autistic -performance on Belief Test depends on Desire Tests -autistic show most profound deficits in understanding emotions caused by beliefs rather than emotions caused by situations and desires -results are consistent with theory that autistic have deficit forming and using metarepresentations -may be a developmental sequence between understanding desires and understanding beliefs -autistic who passed Belief Test had higher CA and MA than those who failed -therefore may be considerable delay in development of theory of mind and it appears in brighter autistic people only in teenage years.

APPENDIX B
SCORING FORMS

Memory Test Scoring Form

Name: ______________

Object	Trials		
	1	2	3
red lego			
green elephant			
blue dog			

Legend

+= correct color identification

- = incorrect color identification

Appearance-Reality Tasks Scoring Form

Name: ____________

Type of Task: Color

		Pretest						Type of Error		
Date	Task	Color	Alternative Format	"A" Question	"R" Question	Pass	Fail	Phenomenist	Realist	Other

Appearance-Reality Tasks Scoring Form

Name: ______________

Type of Task: Size

		Pretest							Type of Error	
Date	Task	Size Discrim.	Alternative Format	"A" Question	"R" Question	Pass	Fail	Phenomenist	Realist	Other

Appearance-Reality Tasks Scoring Form

Name: ______________

Type of Task: Object Identity

Orange: ______ Rock: ______

	Probe Questions						Type of Error		
Date	Realism	Probe	"A" Question	"R" Question	Pass	Fail	Phenomenist	Realist	Other

APPENDIX C

INSTRUCTION SCRIPTS FOR COLOR AND SIZE TASKS

Instruction Script for Color Tasks

Koala Bear and Spaghetti container with red filter
pink eraser/blue filter
white paper/green filter

NB: TEST FIRST BEFORE INSTRUCTING

Complete Modelling Phase

E (Experimenter): **We're going to use this** (bear) (touch yellow bear), **and this** (touch container with filter). **I'm going to show you how things can *look* one way but *really* be something else. What am I going to show you? Listen. How things can *look* one way but *really* be something else. Are you ready? Watch my finger. Look.** (Move finger down to koala bear). **What color is this?**

S (Subject): Yellow.

E: **That's right. It's yellow. Now watch my finger and listen.** (Put bear completely in container). **What color does it look?**

S: Orange (or red).

E: **That's right. It looks orange. Now watch my finger.** (Slides bear 1/4 way under red filter). **Here** (touch part under filter) **it *looks* orange, but** (remove from container) **it's *really* yellow.** (Repeat). **Watch my finger** (put bear 1/4 in). **It *looks* orange, but** (remove bear) **it's *really* yellow.** (E puts bear 3/4 way under filter in container). **Watch my finger.** (Touch part under filter). **It *looks* orange, but** (remove) **it's *really* yellow. Watch my finger** (put 3/4 way under filter). **It *looks* orange, but** (remove from filter) **it's *really* yellow** (E repeats procedure with bear completely under filter in container and remove bear from container to show its real color). (E praises subject for specific things throughout phase, e.g., for looking at finger).

Model and Test Phase

E: **Let's do it again! Watch my finger. Look here.** (Touch bear). **This is yellow. What color is this?**

S: Yellow.

E: (Place bear 1/4 way under filter in container). **Now, watch my finger** (touch part under filter). **It *looks* orange. What color does this** (point) ***look*?**

S: Orange.

E: **Watch my finger. But,** (remove from container) **it is *really* yellow.** (Touch bear). **What color is it really?**

S: Yellow.

E: (Place bear 3/4 way under filter in container). **Watch my finger.** (Touch part under filter). **It *looks* orange.** (Touch part under filter). **What color does it *look*?**

S: Orange.

E: **Watch my finger. But** (remove from container), **it is *really* yellow.** (Touch bear). **What color is it really?**

S: Yellow.

E: (Repeat procedure with bear completely under filter in container). **You followed my finger very well. That's terrific.**

Testing Phase

E: **We're going to do it one more time and now it's your turn. Watch my finger carefully! Where is it? What color is this? It is _______.** (Wait for subject to respond using time delay. Give the proper label after successively longer periods of time. Reinforce student when you give correct response and reinforce even more when correct response is supplied by student).

S: Yellow.

E: **Good! It's yellow** (put bear 1/4 way under filter in container). **Watch my finger** (Touch part under filter). **What color does it *look*? It *looks* _____________.** (Use time delay).

S: Orange.

E: **Yes. It looks orange. But, watch my finger** (remove bear from under filter), **what color is it *really*? It is *really* ___________.** (Use time delay).

S: Yellow.

E: (Slide bear 3/4 way under filter in container. Touch part under filter). **What color does it *look*? It *looks* ____________.** (Use time delay).

S: Orange.

E: **But, watch my finger** (remove bear from under filter), **what color is it *really*? It is *really* ___________.** (Use time delay)

S: Yellow.

E: (Repeat procedure with bear completely under filter in container). **You answered those questions really well! Good work!**

Demonstration of Concept Understanding

E: **We know that this** (touch bear) **is *really* yellow. Can you show me how to make it look orange** (hold filter in S's direction)?

S: (Places bear in container under filter).

E: **Good for you. You made it look ____.**

S: Orange.

E: **But what color is it *really*?**

S: Yellow.

E: **Show me how you know it is really yellow.**

S: (Either removes bear from container or verbally indicates that removing bear would reveal the bear's true color).

E: (Praise S for their performance and then move to second example).

NB: alternate true and apparent color randomly so pattern of responding doesn't develop.

Instruction Script for Size Tasks

Dalmations
Plugs
Washers

NB: TEST FIRST BEFORE INSTRUCTING

Complete Modelling Phase

E (Experimenter): **We're going to use these** (touch dalmations), **and this** (touch plexiglass). **I'm going to show you how things can *look* one way but *really* be something else. What am I going to show you? Listen. How things can *look* one way but *really* be something else. Are you ready? Watch my finger. Look.** (Move finger from in front of subject to dalmations on poster board sitting on easel). **Now listen. Which one is bigger?**

S: (Points to, or accurately verbalizes which dog is bigger). **This one** (correctly points to dog on right). (E corrects if necessary and repeats question).

E: **Good for you. You showed me which one is bigger. Listen carefully now.** (Place plexiglass over objects making smaller dog appear bigger than the objectively bigger dog). **Here** (point to magnified dog) **this one *looks* bigger than this one** (point to dog not magnified), **but** (remove plexiglass) **this one** (point to larger dog) **is *really* bigger. Watch again.** (Place plexiglass over dogs again). **This one *looks* bigger** (point to magnified dog), **but,** (remove plexiglass) **this one** (point to larger dog) **is *really* bigger**. (E praises subject for specific things throughout phase, e.g., for looking at finger, looking through plexiglass etc.).

Model and Test Phase

E: **Let's do it again! Watch my finger. Look here** (touch card with dogs). **Which one is bigger?**

S: This one (points to correct dog).

E: **That's right. This one** (point to bigger dog) **is bigger. Now, watch my finger** (place plexiglass over dogs and point to magnified dog). **Right now, this one *looks* bigger than this one** (point to non-magnified dog). **How does this one** (point to dog under magnifying glass) ***look***?

S: Bigger.

E: **Yes, this one *looks* bigger. But** (remove plexiglass), **this one** (point to larger dog) **is *really* bigger. Which one is *really* bigger?**

S: This one (points to correct dog).

E: **Good watching. Let's do it this way one more time.** (Repeat procedure).

Testing Phase

E: **This time it's going to be your turn to tell me which one *looks* bigger and which one is *really* bigger. Are you ready? Watch carefully! Where are they? Which one is bigger?**

S: This one.

E: (Place plexiglass over dogs). **Which one looks bigger?** (Wait for subject to respond using time delay. Give answer after successively longer periods of time. Reinforce student when you give correct response and reinforce even more when correct response is supplied by student).

S: This one.

E: **Yes. This one *looks* bigger. But, which one is *really* bigger?** (Use time delay).

S: This one (points to correct dog).

E: **Good work. O.K. We're going to do it this way one more time, and you're going to give me the answers again. Are you ready?** (Repeat procedure using time delay to allow the S to provide the answers).

Demonstration of Concept Understanding

E: **We know that this one** (point to bigger dog) **is *really* bigger than this one** (point to smaller dog). **Can you show me how to make *this* one** (point to smaller dog) ***look* bigger than this one** (point to bigger dog)? (Point to plexiglass).

S: (Places plexiglass over objects so that smaller dog is magnified).

E: **Good for you. You made this one** (point to magnified dog) ***look* bigger. But which one is *really* bigger?**

S: (Points to nonmagnified dog).

E: **Show me how you know that this one** (point to nonmagnified dog) **is *really* bigger.**

S: (Either removes plexiglass and verbalizes how it is possible to see that the bigger one is *really* bigger, or removes plexiglass to prove which one is *really* bigger).

E: (Praise S for their performance throughout phase and move to second example).

APPENDIX D
INSTRUCTIONAL FORMAT ANALYSIS

Instructional Format Analysis

Model and Test Phase of Size (Buttons) Task (Phase 2)

Script	Function and Task Analysis
	-general necessary preskills-able to follow finger to see which object is indicated -able to listen to experimenter (attention skills) -able to discriminate size -able to identify objects -able to answer questions -knowledge of terms "looks" and "really"
Let's do it again!	-motivation
Watch my finger.	-focussing prompt
Look here (touch card with buttons).	-focussing prompt
Which one is bigger?	-determine size discrimination ability
That's right.(points to bigger button)	-reinforcement of answer
This one is bigger.	-confirmation and reinforcement of answer
Now, watch.	-focussing prompt
(Place plexiglass over objects.)	-demonstration of how smaller object can be made to look larger than larger object
Right now, this one *looks* bigger than this one (point to magnified button).	-verbalizing appearance -modelling of correct answer -focussing prompt
How does this one *look*? (point to magnified button)	-testing attention and ability to follow the concept presentation -focussing prompt
Yes, this one *looks* bigger.	-reinforcement of answer
But (remove plexiglass) this one (point to larger button) is *really* bigger.	-concrete visual demonstration of concept -pointing serves as focussing prompt
Which one is *really* bigger?	-testing attention and ability to follow the concept presentation
Good watching.	-praise
Let's do it this way one more time.	-reinforcement of concept

APPENDIX E
INFORMATION AND CONSENT FORMS

University of Alberta

Department of Educational Psychology

I am a doctoral student in special education currently conducting a study on autism and "theory of mind". You have received this information because either the Edmonton Autism Society, or your child's school, thought that you might be interested in participating in the study and that your child who has been diagnosed as having autism might fit my requirements. This letter will introduce my topic of study and research plan to you.

"Theory of mind" refers to the ability that young children develop, that enables them to understand and predict the behavior of others, and to recognize and understand the existence of mental states, beliefs, intentions and desires in both themselves and others. A number of studies have shown that a majority of children diagnosed as autistic appear to experience a significant delay in the development of this ability. This delay has been blamed for a number of other observed difficulties including the lack of pretend play, and the social impairments that children with autism tend to have. Although a number of researchers in Britain have consistently found conceptual perspective-taking problems in children with autism, no study investigating the possibility of teaching theory of mind has been attempted. The aim of my study is to see if one aspect of theory of mind can be taught to children with autism who have not already developed this ability.

For my study I plan to use four children as participants who have been diagnosed as having autism. The children that would be in the study will probably be between about 8 and 12 or 13 year of age, and will need to meet certain standards in two standardized tests (the Peabody Picture Vocabulary Test-Revised and the Leiter International Performance Scale), and three theory of mind tasks that I have developed based on those used in previously published studies. I will be under the supervision of my dissertation co-supervisors, Dr. D. Baine, and Dr. R. Mulcahy, Department of Educational Psychology (492-5245).

The initial tasks will be given in two sessions probably lasting for a total of about 1 1/2 hours. Once four children are selected I will be able to begin my study. The gathering of data will involve the administration of different tasks (about 25 minutes maximum) on a daily basis prior to instruction. Once the actual instruction is initiated, I hope to be able to work with your child for two, half-hour periods per day, five days a week up to a maximum of 20 instructional sessions. After the completion of instruction, I will again administer the same tasks as those I administered at the outset (about a half-hour maximum) to see if any gains that may have occurred have been maintained.

If you are interested in the study, there are a number of points of which you should be aware before consenting to the study:

1) I will need access to either medical or school records that document your child's diagnosis and how that diagnosis was obtained (e.g. DSM-IIIR, CARS, etc.)
2) The sessions with your child will be videotaped, but will be viewed only by myself or my supervisory committee unless specific permission is granted by you to show portions of tapes for educational purposes (i.e. conferences). If the opportunity to present my findings at a conference does arise, any portion of any tape that I may consider showing would be available for you to preview for your consent.
3) Participation in the study will necessitate your child being removed from his/her classroom for the duration of all sessions. However, I believe that even if your child does not learn the specific theory of mind tasks, s/he will benefit from the

additional one-on-one instructional time emphasizing attention, eye contact, and responding.

4) You will be absolutely free to withdraw your child from participation in the study at any time.

Should you agree to have your child participate in the study, (and should your child meet the necessary pre-requisites to be included in the study), complete confidentiality and anonymity will be maintained through the use of pseudonyms in any written material. Any and all material pertaining to your child will be available to you throughout the study. As well, a copy of the final dissertation will be available to you upon request.

If, upon reading this letter, you are either interested in having your child participate in my study, or would like additional information, I encourage you to phone me as soon as possible at either 439-9503 (res.), or 492-5245 (university). I would be pleased to provide further details or discuss any concerns you might have concerning your child's potential participation in my study.

Thank you very much for your consideration of my study. I realize that it will be time-consuming for your child to participate, but I feel that even if the primary aim is not met, your child will benefit from the additional individual instruction. I also feel that some very interesting findings will result. I look forward to hearing from you shortly.

Yours sincerely,

Liz Starr, M. Ed.

University of Alberta

Department of Educational Psychology

INFORMED CONSENT FORM

Project Title: Autism and Theory of Mind: Teaching the Appearance-Reality Distinction.

Investigator: Elizabeth M. Starr, M. Ed.

In order to allow your child to participate in my study I request that you read the accompanying information and return the signed form below as soon as possible.

If your child is selected for the study, I will work with him/her twice a day, five days a week, up to a maximum of 20 sessions. Your child's identity will remain confidential at all times, with pseudonyms being used in describing the data. Sessions will be videotaped, but will be viewed only by myself or my dissertation committee, unless specific permission to show portions of the tapes for educational purposes is obtained from you.

If your child is not selected for inclusion in the study, all videotapes of your child that may have been taken during the selection process will be destroyed. Your child's identity will likewise remain confidential.

If you have any questions about the study or if you would like any information at any time before, during, or after the study, you are encouraged to contact me or my supervisors (Dr. David Baine/Dr. Robert Mulcahy, Department of Educational Psychology, 492-5245). As well, you may withdraw your child from the study at any time.

Should you require additional information, or would like to discuss the study in further detail, please do not hesitate to call me at 439-9503 (res.) or at the university at 492-5245 where a message can be left.

This is to certify that I have read and understood the accompanying information and agree to allow my child to participate in Elizabeth Starr's study. I understand that participation will necessitate the administration of the Peabody Picture Vocabulary Test-Revised and the Leiter International Performance Scale. I further understand that even though my child may be administered the inclusion tests, s/he may not be selected for the study. I am aware that sessions will be videotaped but that the name of my child, and all other identifying information will be kept confidential.

Name (please print): ______________________________

Signature of parent or guardian: ______________________ Phone: __________

Signature of researcher: ____________________________

Date: ______________

University of Alberta

Department of Educational Psychology

Autism and Theory of Mind: Teaching the Appearance-Reality Distinction

CONSENT FOR OBTAINING CONFIDENTIAL INFORMATION

By this written consent I, ______________________________ authorize Elizabeth Starr, M. Ed., the principal investigator of the above named doctoral study, to have access to school or medical records pertaining to my child's ([name of child] ________________ diagnosis of autism.

I understand that the University of Alberta is a research institution, and I authorize this information to be used, where appropriate, for research purposes under the direction of a University academic staff member (Dr. D. Baine/Dr. R. Mulcahy, Department of Educational Psychology, 492-5245). I also realize that this information may be published in such a manner that my child's identity will be protected (i.e., through use of a pseudonym).

I understand that I am free to withdraw this consent at any time during the data collection period of the study This consent will cover the period from the present until the date of the principal investigator's final dissertation defense.

This form, or photocopy thereof are equally valid.

Signature of parent or guardian: __________________________

Date: _______________________

APPENDIX F

Session Numbers and Dates

Session Numbers and Dates

Date	Session Number
April 15,1991	1
April 16, 1991	2
No data collected	3
April 18, 1991	4
April 19, 1991	5
April 22, 1991	6
April 23, 1991	7
No data collected	8
April 25, 1991	9
April 26, 1991	10
April 29, 1991	11
April 30, 1991	12
May 1, 1991	13
May 2, 1991	14
No data collected	15
May 6, 1991	16
May 7, 1991	17
May 8, 1991	18
May 9, 1991	19
May 10, 1991	20
May 13, 1991	21
May 14, 1991	22
May 15, 1991	23
May 16, 1991	24
May 17, 1991	25
No data collected	26
May 21, 1991	27
May 22, 1991	28
May 23, 1991	29
May 24, 1991	30
May 27, 1991	31

May 28, 1991	32
May 29, 1991	33